

DEX-0113

BQ

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A2

- (51) International Patent Classification⁷: C07K 14/00 [DE/DE]; Grosse Lachstrasse 30a, 69207 Sandhausen (DE).
- (21) International Application Number: PCT/IB00/01496
- (22) International Filing Date: 18 August 2000 (18.08.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/149,499 18 August 1999 (18.08.1999) US
60/156,503 28 September 1999 (28.09.1999) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/156,503 (CIP)
Filed on 18 August 1999 (18.08.1999)
- (71) Applicant (for all designated States except US): GERMAN HUMAN GENOME PROJECT [DE/DE]; Fraunhofer Patentstelle, Leonrodstrasse 68, 80636 Munich (DE).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): WIEMANN, Stefan
- (74) Agent: CARPMAELS & RANSFORD; 43 Bloomsbury Square, London WC1A 2RA (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— Without international search report and to be republished upon receipt of that report.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/12659 A2

BEST AVAILABLE COPY

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.

HUMAN DNA SEQUENCES

Background of the Invention

Current methods for testing pharmacological substances rely on a three-stage testing approach to drug development. First, candidate compounds are typically screened in some sort of *in vitro* system, like inhibition of cancer cell growth. Candidates are then tested in an animal model, as a first approximation of systemic effects, including efficacy and toxicity. Compounds that still show promise after these initial *in vivo* screens, finally are tested in humans. Again, human testing typically occurs in three phases: toxicity; preliminary efficacy; and efficacy. The entire process can take more than a decade and cost hundreds of millions of dollars. Aside from the monetary costs and protracted time scale, moreover, current testing regimes waste the lives of countless laboratory animals and needlessly endanger the lives of human subjects.

A need exists, therefore, for more sophisticated drug screening techniques that can be done rapidly *in vitro*. These screening techniques ideally will be reflective of systemic and/or organ-specific responses, so that they provide a reliable indicator of action in a human body. Current techniques, however, tend to utilize only a single or limited number of markers, thus answering only very simple questions that are of questionable medical import. For example, a typical *in vitro* assay may ask whether a lead compound binds a particular receptor, which has been implicated in a certain disorder. It is presumed that such binding is indicative of therapeutic usefulness, but it does not even purport to address systemic effects.

Not only are screening techniques for efficacy inadequate, the available toxicity screens likewise are inadequate. Toxicity, on a first level, is usually measured by animal testing. Aside from the complications related to *in vivo* versus *in vitro* testing, such screens are insufficient because of differences in metabolism, uptake, etc., relative to humans. Thus, improved methods would be not only be *in vitro*-based, they would also be more "human."

With the increasing miniaturization of screening assays and the growing availability of targets for pharmaceutical intervention, there is increasing interest in developing arrays containing large numbers of these targets that can be assayed simultaneously. If such an

array contains a large enough population of targets, it can be used to essentially mimic the systemic response. In other words, the array becomes an *in vitro* surrogate for the human body. The more refined the array, the more accurate the predictive capability. In theory, an array could be constructed that can detect all of the known human expression products simultaneously, thereby, providing a very reliable indicator of the human response to a given compound. These arrays offer advantages over the present *in vitro* screening systems in that they can assay large numbers of responses simultaneously. They are superior to animal testing because they are more "human" and, thus, more predictive of human responses.

In order to construct such arrays, however, the field is in need of further human targets. Advantageously, such targets will be provided with additional physiologically relevant information, such as whether the target is expressed in a particular tissue and whether it is related to a known functional class of targets. In this way, the artisan can focus as needed, for example, on tissue-specific effects or target class-specific effects, thereby providing information useful in evaluating efficacy and/or toxicity.

In addition to a need for pharmacological screening targets, there is a need for further pharmacological substances. These substances can be used in the formulation of medicinal compositions and in treating a wide variety of disorders.

The present invention responds to the aforementioned and other needs in the field by providing a population of novel targets useful, *inter alia*, in the profiling and medicinal contexts described above.

Summary of the Invention

It is an object of the invention, therefore, to provide a set of human cDNA clones. Further to this object, the invention provides sequences of human cDNA clones that were isolated from libraries generated from different human tissues.

It is another object of the invention to provide assemblages of targets useful in profiling matrices for screening pharmacological test compounds. According to this object, assemblages comprising different populations of human nucleic acids, proteins and antibodies are provided. In different embodiments, cDNA library-specific assemblages and target-family-specific targets are provided.

It is a further object of the invention to provide a database of human nucleotide and protein sequences. Further to this object, novel human nucleotide and protein sequences are provided in electronic form. In one embodiment, one or more of these sequences is provided in a searchable database.

It is still another object of the invention to provide biologically active target molecules useful in treating or detecting human disorders. Further to this object, the invention provides nucleic acid and protein molecules that have the capacity to affect disease etiology or symptoms or correlate with known disease states. Also further to this object, a database is provided which comprises the disclosed molecules in electronic form.

It is still a further object of the invention to provide polypeptides encoded by the human cDNA clones disclosed herein. Further to this object, the invention provides antibodies and fragments thereof that are capable of binding to a specific portion of these polypeptides.

It is yet another object of the invention to provide pharmaceutical compositions which comprise an effective amount of a pharmaceutical agent, wherein the pharmaceutical agent is selected from the group consisting of one or more polypeptides contemplated by the invention, variants or functional derivatives thereof, and antibodies thereto; and a physiologically acceptable carrier or excipient.

It is still another object of the invention to provide expression vectors comprising one or more human cDNA clones disclosed herein or fragments thereof; and optionally a promoter operably linked to the cDNA clone or fragment thereof. Further to this object, the invention provides methodology for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by a human cDNA clone disclosed herein.

Detailed Description

The invention results from a need in the art for new human nucleic acids and proteins. This need arises in several contexts. First, there is a need to identify targets for therapeutic intervention. Second, there is a need to identify molecules that may be adversely affected in a therapeutic context, thereby resulting in toxicity. Knowledge of these molecules will aid in

the design of new medicaments with enhanced efficacy and decreased toxicity. Finally, the need encompasses human nucleic acids and proteins that have medicinal applicability in their own right.

In view of these needs, the present inventors set out to isolate and sequence human cDNAs from tissue-specific libraries. In this way, they represent subsets of molecules likely to be targets for therapeutic intervention or for avoiding toxicity. In addition, the inventors divided the molecules into various sub-categories, based on suspected functionality, structural similarity etc, which are of interest from a pharmacological perspective. These molecules are disclosed in provisional application serial nos. 60/149,499 and 60/156,503, filed August 18, 1999, and September 28, 1999, respectively, both of which are hereby incorporated by reference in their entirety.

GENERAL DESCRIPTION OF THE INVENTIVE MOLECULES

The present invention provides novel polynucleotide molecules that, in some instances, have similarities with known molecules. The inventive DNAs were cloned from five different human cDNA libraries. In addition to these DNA molecules, the invention provides their protein translations and antibodies derived from them. The inventive DNA and protein sequences are show individually, below. The inventive nucleic acids also include the complements of these DNA sequences, as well as their RNA counterparts. Methods of producing the molecules also are provided. Further, the invention provides methods for detecting all or part of the molecules and of detecting polynucleotides encoding all or part of the molecules.

The inventive molecules derive from five cDNA libraries: human fetal brain; human fetal kidney; human mammary carcinoma; human testis; and human uterus. For convenience, each sequence bears a designation that indicates from which library it is derived. In particular, these designations are: "hfpbr" for human fetal brain; "hfkf" for human fetal kidney; "hmcfc" for human mammary carcinoma; "htes" for human testis; and "hute" for human uterus. The individual libraries were constructed and screened as described below in the examples.

The protein and DNA molecules of the invention are variously described herein as "target" molecules or "inventive" molecules. The sequences and other information pertinent to the nucleic acid and protein molecules of the invention are shown, below.

Interpreting the data disclosed with the Table and cDNA sequences, below:

The table and data below provide the coding sequences of the inventive cDNAs as well as the protein sequences and other useful information, as set out below.

Grouping

The clones were assigned to the following fourteen functional and/or tissue-derived groups:

1. Cell Cycle
2. Cell Structure and Motility
3. Differentiation/Development
4. Intracellular Transport and Trafficking
5. Metabolism
6. Nucleic Acid Management
7. Signal Transduction
8. Transmembrane Protein
9. Transcription Factors
10. Brain derived
11. Kidney derived
12. Mammary Carcinoma derived
13. Testes derived
14. Uterus derived

Description of Clone Files

The individual clone files are structured in the same pattern. The Sections are separated by paragraphs.

1. Clone Name

The clone names are deciphered with reference to the following example:

DKFZphfkd2_24e23, wherein the code represents:

- producer of library ("DKFZ") (for convenience, this reference may be eliminated)
- a "p" for "plasmid cDNA library" (for convenience, this reference may be eliminated)
- library name (e.g. hfbr = human fetal brain; hfkd = human fetal kidney; hmcfc = human mammary carcinoma; htes = human testes; hute = human uterus)
- an underscore ("_") to separate library information from plate information
- plate number (e.g. "16")
- plate coordinates (letter first; e.g. "f14")

2. Group

3. Introduction

short review of the similarities, function of the protein and possible applications

4. Short Information

specifications about the cDNA (who sequenced, completeness of the cDNA, similarity, who sequenced, chromosomal localisation, length of cDNA, localisation of poly A tail and polyadenylation signal)

5. cDNA-Sequence

6. BLASTn Results

search results of blasting the cDNA sequence against all public databases

7. Medline Entries

information about genes/proteins similar to the novel cDNA (if available)

8. Putative Encoded Protein Information

specifications about the encoded protein (ORF: length and localisation of the reading frame)

9. Protein Sequence

10. BLASTp Results

search results of blasting the protein sequence against all public databases

11. Pedant Information

output of fully automated annotation: summarises peptide information, homologues, patterns as follows:

[Length]

- length of the protein = number of amino acid residues

[MW]

- molecular weight of the protein

[pI]

- isoelectric point

[HOMOL]

- shows protein with closest similarity to the cDNA-encoded protein

[FUNCAT]

- functional information according to a catalogue developed by Munich

Information center for Protein Sequences (MIPS)

[BLOCKS]

- Blocks are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. The blocks for the Blocks Database are made automatically by looking for the most highly conserved regions in groups of proteins documented in the Prosite Database. The Prosite pattern for a protein group is not used in any way to make the Blocks Database and the pattern may or may not be contained in one of the blocks representing a group. These blocks are then calibrated against the SWISS-PROT database to obtain a measure of the chance distribution of matches. It is these calibrated blocks that make up the Blocks Database. The WWW versions of the Prosite and SWISS-PROT Databases that are used on this server are located at the ExPASy World Wide Web (WWW) Molecular Biology Server of the Geneva University Hospital and the University of Geneva. World Wide Web URL http://blocks.fhcrc.org/blocks/about_blocks.html/ is the entry point to the database.

- here Blocks segments found in the analysed protein sequences are displayed

[SCOP]

Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a common evolutionary origin. The scop database provides a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known, including all entries in Brookhaven National Laboratory's Protein Data Bank (PDB). It is available as a set of tightly linked hypertext documents which make the large database comprehensible and accessible. In addition, the hypertext pages offer a panoply of representations of proteins, including links to PDB entries, sequences, references, images and interactive display systems. World Wide Web URL <http://scop.mrc-lmb.cam.ac.uk/scop/> is the

entry point to the database. Existing automatic sequence and structure comparison tools cannot identify all structural and evolutionary relationships between proteins. The scop classification of proteins has been constructed manually by visual inspection and comparison of structures, but with the assistance of tools to make the task manageable and help provide generality. Proteins are classified to reflect both structural and evolutionary relatedness. Many levels exist in the hierarchy, but the principal levels are family, superfamily and fold. The exact position of boundaries between these levels are to some degree subjective. Scop evolutionary classification is generally conservative: where any doubt about relatedness exists, we made new divisions at the family and superfamily levels.

- - here SCOPE segments found in the analysed protein sequences are displayed

[EC]

ENZYME is a repository of information relative to the nomenclature of enzymes. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) and it describes each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided. World Wide Web URL <http://www.expasy.ch/enzyme/> is the entry point to the database.

- here EC-number and name of enzymes with similarity to the analysed protein sequences are displayed

[PIRKW]

- functional information according to the Protein Information Resource (PIR) database catalogue developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[SUPFAM]

- information according to the Protein Information Resource (PIR) database catalogue of protein superfamilies developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[PROSITE]

please refer to 12. PROSITE Motifs

[PFAM]

please refer to 13. PFAM Motifs

[KW]

- overall 2dimensional folding information

- 3D indicates that the proteins is similar to a protein of which a 3 dimensional structure is known

- overall structural information

[]

The last PEDANT-block depicts information about the folding structure of the protein generated by PREDATOR. PREDATOR is a secondary structure prediction program. It takes as input a single protein sequence to be predicted and can optimally use a set of unaligned sequences as additional information to predict the query sequence. The mean prediction accuracy of PREDATOR is 68% for a single sequence and 75% for a set of related sequences. PREDATOR does not use multiple sequence alignment. Instead, it relies on careful pairwise local alignments of the sequences in the set with the query sequence to be predicted.

World Wide Web URL http://www.embl-heidelberg.de/argos/predator/predator_info.html is the entry point to the database.

- H = helix, E = extended or sheet, _ = coil, T = transmembrane, B = beta

- x indicates a low-complexity region with repeat-like structure which is omitted in all BLAST searches

12. PROSITE Motifs

PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs. World Wide Web URL <http://www.expasy.ch/prosite/> is the entry point to the database. A description of the prosite consensus patterns is also provided, below.

13. PFAM Motifs

PFAM (protein families) is a large collection of multiple sequence alignments and hidden

Markov models covering many common protein domains. World Wide Web URL

<http://www.sanger.ac.uk/Pfam/> is the entry point to the database.

Deposit of Clones

Clones were deposited as a pool with the American Type Culture Collection under accession number _____, from which each clone comprising a particular polynucleotide is obtainable. Each clone has been transfected into separate bacterial cells (*E. coli*) in this composite deposit.

The clones may also be obtained from the Resource Center of the German Human Genome Project (Heubner Weg 6, 14059 Berlin, GERMANY). The Resource Center library numbers are slightly different than those presented here, but may be readily obtained by the following key or with the assistance of Resource Center personnel.

The library name becomes a number: brain (hfbr2) becomes 564; kidney (hfkd2) becomes 566; mammary carcinoma (hmcfl) becomes 727; testis (htes3) becomes 434; and uterus (hute1) becomes 586. Next, the plate number is converted to two digits (e.g., "2" becomes "02") and is moved behind the plate coordinate, and the underscore is dropped. The following examples are helpful:

<u>Listed Number</u>	<u>Resource Center Number</u>
DKFZphfbr2_16f21	DKFZp564F2116
DKFZphfkd2_1j9	DKFZp566J091
DKFZphmcf1_1c23	DKFZp727C231
DKFZphtes3_14g5	DKFZp434G0514
DKFZphute1_17k7	DKFZp586K0717

The libraries were constructed using two commercially available vectors. The brain (hfbr2 designations) and kidney (hfkd2 designations) libraries utilize pAMP 1 from Life Technologies and are maintained in XL-2Blue (Stratagene); the uterus (hute1), testes (htes3) and mammary carcinoma (hmcfl) libraries are constructed in pSPORT1, also from Life Technologies, and are maintained in DH10B (LifeTechnologies). In addition to the following techniques, consultation with the commercial literature available on these clones will make evident all of the housekeeping techniques needed to propagate and isolate the individual constructs. All inserts may be excised with a NotI/SalI digestion. Alternatively, universal primers, flanking the cloning region, may be used to amplify the inserts using PCR methods.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a combination of those sequences. Methods of probe design are presented below.

Oligonucleotide probes may be labeled with γ - ^{32}P ATP (specific activity 6000 Ci/mmole) and T4 polynucleotide kinase using commonly employed techniques for labeling oligonucleotides. Other, non-radioactive labeling techniques can also be used.

Unincorporated label typically is removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe can be quantified by measurement in a scintillation counter. Preferably, specific activity of the resulting probe generally should be approximately 4×10^6 dmp/pmole.

The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 μl of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 50 - 100 $\mu\text{g/ml}$ (for XL-2Blue strains 25 $\mu\text{g/ml}$ tetracycline should also be used). The culture should preferably be grown to saturation at 37°C ., and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing ampicillin at 100 $\mu\text{g/ml}$ (for XL-2Blue strains 25 $\mu\text{g/ml}$ tetracycline should also be used) and agar at 1.5% in a 150 mm petri dish when grown overnight at 37°C . Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them. The filter is then preferably incubated at 65°C . for 1 hour with gentle agitation in 6 x SSC (20 x stock is 175.3 g NaCl/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 $\mu\text{g/ml}$ of yeast RNA, and 10 mM EDTA (approximately 10 mL per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater than or equal to 1×10^6 dpm/mL. The filter is then preferably incubated at 65°C . with gentle agitation overnight. The filter is then preferably washed in 500 mL of 2 x SSC/0.5% SDS at room temperature without agitation, preferably followed by 500 mL of 2 x SSC/0.1% SDS at room

temperature with gentle shaking for 15 minutes. A third wash with 0.1 x SSC/0.5% SDS at 65°C. for 30 minutes to 1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing.

Alternatively, clones may be grown as described above, and PCR used to isolate the insert DNAs. Methods of PCR are described below and are otherwise well known .

ERROR SCREENING

The DNA sequences found herein derive from individual clones, which are publicly available, as noted above. Thus, the skilled artisan will recognize that any specific sequence disclosed herein readily can be screened for errors by resequencing a particular fragment, in both directions (*i.e.*, by sequencing both strands). Alternatively, error screening can be performed by amplifying and/or cloning any of the inventive DNAs, using for example RT-PCR, and sequencing the resulting amplified product. In the event that there is a sequencing error, reference should be made to the deposited clone as the correct sequence.

USES AND BIOLOGICAL ACTIVITIES OF THE INVENTIVE MOLECULES

The inventive molecules and their derivatives are susceptible to a wide variety of uses, based on functional and/or structural properties. The skilled worker will appreciate, based on the biological activities detailed below, and discussed with regard to the individual sequences disclosed below, that the inventive molecules will find usefulness in numerous therapeutic and diagnostic applications.

The DNA molecules, especially the potassium salts thereof, can be used as fertilizer supplements due to their high nitrogen and phosphorus contents. Since the DNAs are of defined length, they are also useful in gel electrophoresis as molecular weight markers. Due to their similarity with known molecules, certain of the DNA molecules and their variants and derivatives may be used in any number of different diagnostic procedures and therapeutic applications. They may also be used to make the encoded proteins.

The proteins themselves have many possible uses. They may be used as a nutritional supplement for humans, animals and even for laboratory use as, for example, medium for bacterial cultures. Moreover, since the proteins are of defined, known sizes, they may be used as molecular weight markers for gel electrophoresis and gel filtration. Because they are of defined sequences, they also have use in microsequencing and protein fingerprinting applications.

Expression Profiling Applications

Given their known tissue expression and functional associations, assemblages of the inventive proteins (or corresponding antibodies) and nucleic acids are particularly suited to expression profiling applications. Expression profiling generally entails constructing an array of indicators that signal the presence of a particular RNA or protein expression product. Such arrays can be used to evaluate, for example, pharmacological effectiveness and toxicity. In particular, expression profiles from such arrays can be generated from cells treated with known compounds, having known properties, and these profiles can be compared to profiles of unknowns to evaluate similarities and differences, which can be correlated with efficacy or toxicity.

Additional uses of profiling include diagnosis, tracking development, and ascertaining signaling and metabolic pathways. For examples of references describing profiling and its uses, see Farr *et al.*, U.S. Patent 5,811,231 (1998); Seilhamer *et al.*, U.S. Patent 5,840,484 (1998); Rine *et al.*, U.S. Patent No. 5,777,888 (1998); WO 97/27317; WO 99/05323; WO 99/09218; and WO 99/14369. For a device for implementing such techniques, see Lipshutz *et al.*, U.S. Patent No. 5,856,174 (1999) and Anderson *et al.*, U.S. Patent No. 5,922,591 (1999).

In one embodiment, a subset of the inventive DNAs will be arrayed on a substrate, like a gene chip, a filter or a 96-well plate. Test samples containing cells are maintained in the presence of a label capable of incorporation into nascent mRNA. Samples are treated with test and control compounds, which will induce mRNA expression in the sample, resulting in incorporation of label. Whole mRNA is isolated and applied to the array such that it hybridizes with the DNAs contained therein. After washing, the amount of hybridization is quantified and a profile is generated. These steps are repeated with various control and test compounds, thereby generating a library of profiles, which can be used to ascertain the relationships relevant to pharmacological efficacy or toxicity.

The matrices used in such profiling, however, need not be limited to those utilizing DNAs. Rather, other nucleic acids, like RNAs and protein nucleic acids (PNAs), as well as the inventive proteins and antibodies corresponding to the inventive proteins may also be employed. Hence, for example, antibodies could form the array and the samples could be treated in order to label nascent proteins. Whole proteins then would be isolated and applied to the antibody matrix. Developing the resulting signal would result in a protein expression profile, which is useful in essentially the same manner as the nucleic acid profile. A protein matrix could be used, for example, in evaluating antibody responses to pharmaceutical agents in order to eliminate possible cross-reactivity.

Moreover, where nucleic acids are used in the matrix, it is often beneficial to use variants (as defined below) of the molecules described herein. This can be used to account for genetic variations that are of little or no consequence to the function of the resultant gene product. Hence, they can account for wobble or conservative amino acid variations that do not perturb function, like variations in some of the protein motifs elucidated below. Thus, each position in the matrix can employ multiple nucleic acid probes that account for a series of variants.

Expression profiling may also be done, in another embodiment, using two-dimensional protein gels in which the inventive proteins are detected. The resultant profiles can be used in the same way as described.

Matrices useful for profiling may be constructed based on different criteria. Of course, the more relevant profiles will take into account expression of most human genes, preferably all of them. In certain situations, however, it is advantageous to look at a smaller subset. For example, if one were concerned about fetal neural toxicity, a fetal brain-specific matrix might be chosen. On the other hand, if one were interested in targeting mammary carcinoma tissue, a corresponding matrix could be used. Thus, matrices may be constructed using all of the sequences available from a tissue-specific library.

* * *

The following discussion relates to some of the various functional and structural groupings that would be of interest to the artisan wishing to construct profiling matrices. Of course, the artisan will also recognized that these functional descriptions may find additional applicability in the therapeutic and diagnostic applications discussed below.

Cell Cycle

A proliferating cell must coordinate replication and chromosomal separation to ensure that the genome is replicated completely, and that a single copy is correctly inherited by each daughter cell. The cell cycle is the coordinated series of events that achieves these aims. Many of the key events are initiated by a family of conserved Serine/threonine protein kinases, the cyclin-dependent kinases (CDKs), that are activated by the cyclin family of proteins (cyclins A-H). In turn, the cyclin-CDK complexes are modulated by other protein kinases or phosphatases, and by binding specific inhibitor proteins. The enormous variety of ways in which CDK activity can be regulated allows the cell to respond to internal signals generated by preceding events in the cell cycle and to external growth signals.

The somatic cell cycle is divided into four phases: DNA replication (S phase) and chromosome separation (M phase) are separated by gap phases (G1 and G2). At specific control points the decision to begin the next stage (DNA synthesis or mitosis) is carefully regulated.

Cdc2, the primary kinase, is especially required for the G1-S transition and S phase. Cdc4 and Cdc6 are involved at the restriction point, where the cell can decide to proliferate or arrest (G1 \leftrightarrow G0) and Cdc7 is a CDK activating kinase (CAK) as well as a subunit of TFIIF.

The Cyclin-CDK complexes are regulated in various ways. One is through phosphorylation by CDK activating kinases (CAK), like the Y15 kinase (Wee1) and dephosphorylation by CDK associated phosphatases (CAP), like Cdc25A a member of the Cdc25 family (Cdc25A, B and C).

Another way of regulation occurs through two classes of CDK inhibitors (CKI), the INK4 proteins p15, p16, p18, and p19, who negatively regulate the cyclin D CDK complexes and second the p21 family with p21, p27, and p57.

The cell cycle is also regulated through ubiquitin-mediated proteolysis involving the destruction of both cyclins and CDK inhibitors by the 26S proteasome, that requires an ubiquitin conjugating enzyme (UBC) and an ubiquitin ligase. The instability is conferred by PEST regions (cyclin D and E) or a ten amino acid region in the amino terminus (degradation box) in the A- and B-type cyclins.

All these modifications play an important role for the cellular localization, because only the nuclear CDK-cyclin complexes are functional for cell cycle. During G1 phase of the cell cycle, cyclins A, E and D are synthesized and bind to their cyclin-dependent kinase (CDK) partners. CDK complexes containing cyclins A, E and D1 are then imported into and concentrated within nuclei. Cdk6- cyclin D3 has been localized to both cytoplasmic and nuclear compartments, although only the nuclear complex is active. As cells enter S phase, cyclin A and cyclin E complexes remain within the nucleus, whereas cyclin D1 relocates to the cytoplasm for proteolysis at the onset of S phase. Like Cdk2-cyclin A, Cdc2-cyclin A is nuclear and remains so until it is degraded during mitosis. By contrast, as a result of ongoing nuclear import and more rapid re-export, cyclin B1, which binds to Cdc2 upon synthesis during S phase, is predominantly cytoplasmic. Cdc2-cyclin B2 is also cytoplasmic, although this might occur through anchoring of the complex to some cytoplasmic constituent. At prophase, phosphorylation of cyclin B1 promotes accumulation of Cdc2-cyclin B1 in the nucleus, whereas cyclin B2 remains in the cytoplasm until nuclear envelope breakdown.

Two crucial regulators of Cdc2-cyclin B-Wee1 and Cdc25C exist and are responsible for the G2 to M control point. Wee1 is a nuclear protein throughout the cell cycle, whereas Cdc25C binds to 14-3-3 proteins during interphase and remains predominantly cytoplasmic. In some systems Cdc25C, like cyclin B1, rushes precipitously into the nucleus just before entry into mitosis.

The 110-kDa retinoblastoma (tumor suppressor) protein (RB), a pRB-family member is an important regulator of cell-cycle progression and differentiation. Like the E2F family (E2F1-5) or DP family (DP1-3) of transcription activators, RB suppresses inappropriate proliferation by arresting cells in G1 by repressing the transcription of genes required for the transition into S phase. Before the cell proceeds into S phase, RB becomes phosphorylated at multiple sites by the cyclin dependent protein kinases (CDKs) and loses its transcriptional repressing activity. Phosphorylation of RB during late G1 phase results in the dissociation of the E2F-RB repressor complex which allows S-phase specific genes to be transcribed. Cyclin E is the evolutionary conserved target for E2F and interacts together with CDC2 in late G1.

For a proliferating cell it is vital that only undamaged DNA is replicated because if DNA damage is substantial, its replication can lead to chromosome loss or rearrangement.

Thus, we find a G1 \leftrightarrow S checkpoint in late G1 that requires tumor suppressor p53. A p53-dependent G1 arrest is effected by the cyclin dependent kinase inhibitor p21 through higher expression levels that inhibits almost all cyclin CDK complexes.

The kinase responsible for phosphorylating the unidentified kinetochore component in metaphase may be a member of the MAP kinase family and appears to be the proto oncogene c-MOS, a cytostatic factor (CSF) in meiosis.

Several categories of proteins are coded for by clones of the invention within the overall group of "Cell cycle" and include, among others, the following:

Tumor suppressors (e.g. N33): Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. The N33 gene has been reported by OMIN OMIN (Online Mendelian Inheritance in Man at <http://www.ncbi.nlm.nih.gov/htbin-post/Omin>) to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) prostate cancer suppression (OMIN *601385). Clones in this category include: fbr2_2k14.

C-TAK1 Cdc25c associated protein kinase: Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. Alterations in the gene coding for the above protein kinase has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Pancreatic cancer (OMIN *60278). Clones in this category include: tes3_7j3.

Cell structure and motility

One of the major differences between prokaryotes and eukaryotes is the ability of the eukaryotic cell to adopt very different shapes dependent on its function during the differentiation process. Animal cells vary from being round to extended cylindric forms like motoneurons or muscle cells. In humans, more than 100 different cell types can be distinguished, each having a characteristic shape. The form of a cell often is closely related to

its capacity to move. Some completely differentiated cells like fibroblasts can still change their form actively, thereby migrating. Other cell types serve as motor elements - "macroscopically" like muscle cells or "microscopically" like ciliated epithelia. Such tasks are fulfilled by a big class of proteins; on the one hand responsible for maintenance of cell structure and contacting neighbor cells or the intercellular matrix and on the other hand for cell motility. These topics cannot be regarded separately: The motility apparatus e.g. must be fixed in the cytoskeleton. Three different types of filaments can be distinguished: Actin filaments, tubulin filaments and intermediate filaments, each present in almost all types of cells.

Actin filaments (F-actin) are built up of monomers (G-Actin). In muscle cells, actin, myosin, for both of which several paralogous genes are known, as well as many more proteins are constituents of the contractile apparatus.

The "thin" and "thick filaments" in a muscle cell consist mainly of actin and myosin, respectively.

Several different proteins are responsible for the anchoring of the actin filaments in the Z-disks (e.g. alpha-actinin and desmin) or at the end of the myofibers in the cell membrane.

Troponin I, -C, -T and Tropomyosin - associated with actin - confer the Ca^{++} -dependent triggering of contraction.

Length of the sarcomere is controlled by the giant protein titin.

In smooth muscle, there is no troponin. Contraction activity is controlled by phosphorylation / dephosphorylation of myosin by a specialized kinase instead. Contractile fibers are not organized in sarcomeres.

Apart from contributing to muscle contraction, the actomyosin system is responsible for many other motions at cellular level, e.g. the amoeboid movement of pseudopodia or the fission of cells at the end of mitosis by a contractile ring.

Besides this, actin fibers fulfill structural tasks like maintenance of the shape of stereocilia or microvilli. Here, actin filaments are connected by proteins like fimbrin. But not

only specialized structures like the mentioned ones contain actin fibers. There is a network covering the complete cell volume with F-actin as a major constituent. Whereas the actin filaments in the structures mentioned above are relatively stable, this F-actin is highly dynamic. Management of the network structure and turnover is achieved by connecting proteins like alpha-actinin, fimbrin or filin; turnover is regulated by gelsolin, villin, and different capping- and fragmentation-proteins.

Microtubules are built up of alpha-beta tubulin heterodimers. Turnover of filaments is achieved by building-in and releasing of monomers with different time constant rates at both ends. The resulting cycle is called "treadmilling". Thirteen strings of tubulin duplets build up one subfiber, whereas one fiber contains two or three of those. A complete axoneme consists of 9 radial and 2 central fibers. This "9+2" - structure is the basis both of flagella, their basal bodies and centrioles. In flagella, several additional structures like radial elements exist. Nexin connects the fibers and dyneine is the motor ATPase which shifts the fibers relative to each other. Several genetic diseases like the Kartagener syndrome are caused by deficiencies of distinct proteins in cilia.

Besides this, microtubules are abundant in all types of cells. They are part of a delivery system for organelles, e.g. in the golgi apparatus. A further very important system based on microtubules is the mitotic spindle, it is organized by the centrosomes. Besides many other components, the major part of a centrosome are two centrioles which are built up of nine microtubule-triplets. Most remarkably, new centrioles are not synthesized de novo but generated by duplication of old ones.

Cytoplasmic microtubules are associated with many different proteins. Two major classes are known: The MAPs ("microtubule-associated proteins", with molecular masses between 200 and 300 kD) and the much smaller tau-Proteins with a MW between 60 and 70 kD. These proteins regulate the treadmill-process and the interaction with other structures in the cell.

Besides actin and myosin the so-called intermediate filaments constitute a third class of filaments. In contrast to the former two groups, they do not participate in motility, nor are they dynamic structures subject to a vivid turnover. The most important ones are

neurofilaments (in neurons), keratin filaments (mainly in epithelial cells), and vimentin filaments (in many sorts different cell types).

The biological function of both the cytoskeleton as well as contractile apparatus of a cell does not end at the cell membrane. Cells must be embedded in the extracellular matrix, all cells of a muscle must act as one single mechanical unit and epithelia must resist macroscopic mechanical forces. Hence, cell adhesion and the extracellular matrix are closely connected to the cytoskeleton. Vincullin is one of the proteins which serve as an anchor for intracellular fibers (actin). Different types of desmosomes and tight junctions connect neighbor cells with intercellular fibers. On the inside, cytoplasmic plaques connect them to the cytoskeleton. These structures, on the one hand, serve as mechanical elements whereas gap junctions, on the other hand, connect cells metabolically.

The extracellular matrix consists of a network of proteins, glycoproteins and polysaccharides. Different proteins are present in relation to different mechanical demands: Elastin is found in tissues with high elasticity (lungs, heart) whereas collagen, a more hard-wearing protein, is found in tendons and ligaments. Fibronectin is an extracellular protein highly important for cell adhesion.

Reference: Murray J *et al* (1992): Cell Motil Cytoskeleton 22: 211-223.

Within the overall group of Cell Structure and Motility several categories of proteins are coded for by clones of the invention:

Collagen alpha chain proteins: Proteins with the typical (xxG)_n repeat of collagen proteins and Pfam von Willebrand factor type A domain(s) suggest they are collagen alpha chains. These proteins can find application in modulation of connective tissue, bone and cartilage development and maintainance. OMIN reports collagen alpha chains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Osteogenesis imperfecta, type I (OMIN #166200); 2) Osteogenesis imperfecta congenita (OMIN #166210); 3) Alport Syndrome, X-linked (OMIN #301050); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Ehlers-Danlos Syndrome, Type VII (OMIN #130060); 6) Marfan Syndrome (OMIN #154700); 7) Alport Syndrome, Autosomal Recessive (OMIN #203780); 8) Alpha-2-Deficient Collagen Disease (OMIN 203760); 9) Goodpasture Syndrome (Omin 233450); 10) Osteogenesis Imperfecta,

progressively deforming, with normal sclerae (OMIN #259420); 11)) Ehlers-Danlos Syndrome, Type VII Autosomal Recessive (OMIN *225410); and 12)) Osteogenesis imperfecta, Type IV (OMIN #166220). OMIN reports that von Willebrand factor type A domains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases:: 1) Hemophilia A (OMIN *306700); 2) Von Willebrand Disease (OMIN *193400); 3) Giant Platelet Syndrome (OMIN *231200); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Congenital Thrombotic Diseases due to protein C deficiency (OMIN #176860); 6) Polycystic Kidney Disease 1 (OMIN *601313); 7) Nephrogenic Diabetes Insipidus (OMIN *304800); 8) Factor V Deficiency (OMIN *227400); and 9) Dentatorubral-Pallidolusian Atrophy (Omin *125370). Clones in this category include: fbr2_2b5.

Radial spokehead protein: Radial spokehead proteins, e.g., Chlamydomonas reinhardtii radial spokehead protein of flagella or axoneme and the Strongylocentrotus purpuratus sea urchin spermatozoa protein p63, and human proteins with similarity thereto are important for the maintenance of a planar form of sperm flagellar beating. The human protein(s) can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men (e.g., in sterility). Clones in this category include: tes3_15i5.

Ankyrins: Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus these proteins are involved in coupling of cyto skeleton and cell membrane. OMIN reports that Ankyrins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Hereditary Spherocytosis (OMIN *182900); 2) Hemolytic Poikilocytic Anemia due to reduced ankyrin binding sites (OMIN 141700); 3) Atypical Elliptocytosis (OMIN 225450); 4) Autosomal recessive spherocytosis (OMIN #270970); 5) Werner Syndrome (OMIN *277700); and 6) Rhesus-unlinked type Elliptocytosis (OMIN #130600). Clones in this category include: tes3_18i7.

FGD1-related F-actin binding protein (Farbin/FGD1): FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. (OMIN 305400). Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as

described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an esin yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. Clones in this category include: tes3_72k15.

Paramyosins: Paramyosin is a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*. Clones in this category include: tes3_7b22.

Tuftelin: Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix. The new protein can find application in modulation of tissue-calcification, especially the uterus. As reported by OMIN, tuftelin has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with amelogenesis imperfecta (OMIN *600087). Clones in this category include: ute1_19g22.

Cell Adhesion Regulator (CAR1): CAR1 is involved in the regulation of cell-cell adhesion. OMIN reports the association (as potentially diagnostic, therapeutic, causative, and/or related, etc...) of CAR1 with tumor suppression by the reduction of tumor invasion (OMIN *116935). Clones in this category include: ute1_24j6.

Differentiation/Development

Almost every multicellular organism originates from meiotic cell divisions and the recombination of a paternal and a maternal set of chromosomes. After fertilization of the egg, all cells of a body originate from this one cell. Thus the cells of the developing body are initially genetically alike. But phenotypically they become very different. They are specialized to a certain cell type and arranged in an organized pattern to a certain type of tissue and the whole structure has the well-defined shape of an organ. All these features are determined by the DNA sequence of the genome, which is reproduced in every cell. Each cell acts on the genetic instructions given to a certain time and at a certain place of development and plays its individual part in the multicellular organism. Cell differentiation may be divided into three general steps: cell cycle exit, apoptosis protection and tissue specific gene

expression. These processes are coordinated to provide the final and unique tissue characteristics.

An animal cell that has achieved a certain level of development is said to be determined. This differentiation of a cell may be irreversible and in that case the cell may be renewed only by simple duplication. Other cells are renewed by means of stem cells which are immortal (e.g. stem cells of the bone marrow, epidermal stem cells). The genetic control of development is extensively studied in non-vertebrates and vertebrates. The classical animal model is the fruit fly *Drosophila* and the modern model is the transgenic mouse. Animal transgenesis has proven to be useful for physiological as well as physiopathological studies. Besides the approach based on the random integration of a DNA construct in the mouse genome, gene targeting can be achieved using totipotent embryonic stem cells for targeted transgenesis. Transgenic mice are then derived from the embryonic stem cells. This allows the introduction of null mutations in the genome (so-called knock-out) or the control of the transgene expression by the endogenous regulatory sequence of the gene of interest (so-called knock-in). Mice can be created that express wild-type genes, mutant genes, marker genes or cell lethal genes in a tissue specific manner. These animal models allow to follow changes in tissue and organ development and lead to a better understanding of the cellular function of many genes or to the generation of animal models for human diseases. Fundamental problems in immunology, onset and development of cancer, regulation in fatty acid metabolism, aspects of cardiovascular function, control of the central nervous system development, analysis of reproductive development and function are only some examples of research interests.

The final stage of cell differentiation is growth arrest. In animal tissues with rapid cell turnover terminally differentiated cells undergo programmed cell death. The cells have the ability to kill themselves by activating an intrinsic cell suicide program when they are no longer needed or have become seriously damaged. The execution of this program is termed apoptosis. Apoptosis is of importance for development and homeostasis of animals. The key components of this program have been conserved in evolution from worms (*C. elegans*) to insects (*Drosophila*) to humans. The roles of apoptosis include the sculpting of structures during development, deletion of unneeded cells and tissues, regulation of growth and cell number, and the elimination of abnormal and potentially dangerous cells. In this way

apoptosis provides "quality control mechanism" that limits the accumulation of harmful cells, such as virus-infected cells and tumor cells. On the other hand inappropriate apoptosis is associated with a wide variety of diseases, including AIDS, neuro-degenerative disorders and ischemic stroke. Because it is now clear that apoptosis is a result of an active, gene-directed process, it should be eventually possible to manipulate this form of cell death by developing drugs that interact with its recently identified mechanisms of action. Inducers of cell differentiation, cell cycle arrest and apoptosis might be the novel molecular targets for new anticancer agents in addition to the signaling pathways for growth factors and cytokines.

Proteins, factors, receptors and genes of importance in apoptosis:

Proteases:

- Calpain, an intracellular cysteine protease, exact role unknown.
- Caspase-1 to Caspase-11, a family of proteases synthesized as an inactive proenzyme. Targets of the activated enzymes include: poly(ADP-ribose) polymerase, DNA-dependent protein kinase, U1 ribonucleoprotein, nuclear laminins and cytoskeleton components (actin).
- Granzyme B, a serine protease released by cytotoxic T-cells.

Receptors:

- CD 95 (synonyms: Fas, APO-1), a receptor protein of the TNF-receptor family which includes TNF-R1 and TNF-R2 with the common characteristic of a 70 amino acid cytoplasmic domain.
- FADD (synonym: MORT-1), a cytoplasmic protein
- DR-3 (synonym: APO-3) a member of the TNF-receptor-family
- DR-4 and DR-5

Genes:

- ced-3, ced-4 and ced-9 encode the general apoptotic and antiapoptotic program in *Caenorhabditis elegans*. Apaf-3 is the mammalian homologue of ced-3.

- Bcl-2 / Bcl-xL / Bax / Bcl-xS / Bak: a large gene family that can either inhibit or promote apoptosis.

- Cytokine response modifier A, a cowpox virus gene whose gene product inhibits caspases.

Others:

- Caspase-activated DNase (CAD) and its inhibitor (ICAD), causes DNA fragmentation in the nucleus

- Ceramide, a complex lipid that acts as a second messenger.

- c-Jun N-terminal kinase (JNK) is a proline-directed kinase

- p53 protein, is essential for the induction of apoptosis as a response to chromosomal damage.

- RAIDD, a death signal-transducing protein.

- Receptor interacting protein (RIP) is an accessory protein with a death domain and a serine/threonine kinase activity.

- Sphingomyelinase, an enzyme that hydrolyzes the complex lipid sphingomyelin to ceramide.

- Tumor necrosis factor (TNF) is a type -II membrane protein

- TNF-receptor associated factor (TRAF2), is an accessory protein that can bind to both TNF-R1 and TNF-R2.

Within the overall group of Differentiation/Development, several categories of proteins are coded for by clones of the invention:

Interleukins (e.g. Interleukin-7): Interleukin precursors related to interleukin-7, for example, are expected to act as new growth factors for human B lineage cells. Additionally,

these proteins should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells. These interleukins could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells. (OMIN *146660). Clones in this category include: tes3_35e21.

Testis-specific Y-encoded proteins: The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. Proteins of the TSPY-SET-NAP1L1 family represent proteins closely related to TSPY. These proteins seem to be involved in early spermatogenesis. Clones in this category include: fbr2_2d15.

Intracellular transport and trafficking

Eukaryotic cells rely for their viability on the partitioning of many basic cellular processes into membrane-bounded organelles. These are the nucleus, endoplasmic reticulum (ER), Golgi apparatus, endosomes, lysosomal compartments, mitochondria and peroxisomes. Most molecules destined for the lysosome, cell surface and outside the cell are routed through the ER and Golgi, which together with the vesicular intermediates between them, comprise the secretory pathway (Palade 1975). In the ER and Golgi compartments proteins are sorted, modified and often assembled into complexes *en route* to their final destination. Incorrectly assembled proteins are retained in the ER until they fold correctly or are targeted for degradation. Additional proteins are translocated into and function within the luminal spaces of organelles or are secreted. Thus a large proportion of proteins synthesized require targeting to membranes either for insertion into or transport across them. A major purpose of this is growth. The secretory pathway is dependent on an intact cytoskeleton and also closely linked to general metabolism by affecting ribosome biogenesis (Mizuta and Warner, 1994). A huge number of proteins is required for targeting, translocation and sorting of newly synthesized proteins.

The first step in sorting is the recognition of cis-acting targeting or signal sequences that organelle-targeted proteins contain. This is carried out by cytosolic targeting factors and/or receptors on the membrane to which the protein is targeted. In some cases the primary

sequences are extremely degenerate, with only the overall character being conserved (hydrophobicity for an ER signal sequence, helical amphiphilicity for mitochondrial targeting sequence (Kaiser *et al.*, 1987; Lemire *et al.*, 1989). Following the targeting step, proteins are either inserted into or transported across the membrane (translocated) through a proteinaceous apparatus (termed the translocon). The translocon include or recruit motors to drive the translocation process in the correct direction (Schatz and Dobberstein, 1996).

Defined intracellular protein transport steps:

- ER
 - targeting to the ER
 - translocation into the lumen of the ER, and, depending on the presence of certain signals in the peptide sequence transport through the golgi complex
- Mitochondria
 - targeting
 - translocation
- Peroxisomes
- The general secretory pathway
 - protein modification, assembly and quality control in the ER
 - vesicle-mediated trafficking
 - vesicle docking and fusion
 - transport through the golgi apparatus and sorting at the trans-golgi
 - transport to the cell surface
 - transport routes to the lysosome
- Endocytosis
- Specialized protein transport routes
- Protein export from the cytoplasm

References: Palade, G (1975) Science 189:347-358; Mizuta et al. (1994) Mol Cell Biol 14: 2493-2502; Kaiser *et al.* (1987) Science 235: 312-317; Lemire *et al.* (1989) J Biol Chem 264: 20206-20215; Schatz et al. (1996) Science 271: 1519-1526.

Rab proteins

In eukaryotic cells the compartmentalisation of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins

and other molecules. Trafficking between organelles within the secretory pathway occurs as vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes in vitro, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation

inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle, most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown in vitro to interact with -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca^{2+} -binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca^{2+} influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A

homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn²⁺-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991) *Mol. Cell. Biol.* 11, 872-885; Echard et al. (1998). *Science*. 279, 580-585; Geppert et al. (1998) *Annu. Rev. Neurosci.* 21, 75-95; Guo et al. (1999). *EMBO J.* 18, 1071-1080; Kato et al. (1996) *J. Biol. Chem.* 271, 31775-31778; Novick et al. (1997) *Curr. Opin. Cell Biol.* 9, 496-504; Peterson (1999) *Curr. Biol.* 9, 159-162; Poirier et al. (1998) *Nat. Struct. Biol.* 5, 765-769; Vitale et al. (1998) *EMBO J.* 17, 1941-1951; Wang et al. (1997) *Nature*. 388, 593-598; Yang et al. (1999) *J. Biol. Chem.* 274, 5649-5653.

Within the overall group of Intracellular Transport and Trafficking several categories of proteins are coded for by clones of the invention.

Rab proteins:

Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells. . Clones in this category include: fbr2_2i17, fbr2_3b16.

Rab10 appear concentrated on membranes in the perinuclear region. Rab 10 has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Choroideremia (OMIN *303199); and 2) RETT Syndrome (OMIN 312750). Clones in this category include: fbr2_62119.

In mice, Rab17 shows epithelial cell specificity. Rab 17 is discussed as candidate gene for the mouse mutations ln (leaden), Tw (twirler), and ax (ataxia). Cloned from a brain cDNA library, the new putative Rab-protein is expected to be involved in vesicle trafficking within neuronal cells. These proteins can find application in modulating the transport of vesicles inside neuronal cells, which are essential for development of functional dendritic processes. . . Clones in this category include: fbr2_41m15.

Ankyrin G: The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments. Ankyrin G has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Werner disease (OMIN *277700). Clones in this category include: fkd2_24p5.

Zn-T-transporters: The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide. These proteins can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation. (OMIN *602878, *602095). Clones in this category include: fbr2_62f10.

Metabolism

This group includes proteins which are involved in the uptake and consumption of nutrients, and enzymes which are part of the biochemical pathways for energy metabolism or

which are involved in the supply of building blocks of nucleic acids, proteins (NTPs, dNTPs, amino acids) for DNA/RNA and protein synthesis, and fatty acids (membranes), to allow for the generation of higher order structures. This group constitutes the most important and largest group in prokaryotes and lower eukaryotes. The higher the evolutionary level of an organism is, however, the more other protein classes like 'signal transduction', 'cell cycle' and 'differentiation and development' increase in importance and number of representatives.

Proteins involved in the metabolism of energy and compounds (here: other than nucleic acids or proteins) are usually the products of house keeping genes, they are often constitutively and/or ubiquitously expressed.

Several categories of proteins are coded for by clones of the invention within the overall group of Metabolism:

NAT1, ARD1: In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. These can find application modulating NAT assembly and action and therefore could be important in metabolism of drugs and environmental mutagens.(OMIN *108345). Clones in this category include: fbr2_3g8.

Apolipoprotein E receptor: In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands. These proteins can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins. In normal individuals, chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. In familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants because of a defect in apolipoprotein E. Accumulation of the remnants can result in xanthomatosis and premature coronary and/or peripheral vascular disease. OMIN reports that apolipoprotein has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Familial combined hyperlipidemia (OMIN 144250); and 3) Alzheimer disease. (OMIN #104300). Clones in this category include: fbr2_62017.

Ubiquitin carboxyl-terminal hydrolases: Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. OMIN reports that Ubiquitin-specific proteases have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Lung carcinoma (OMIN *603486); 2) x-linked retinal diseases (OMIN *300050); 3) oncogenesis (OMIN *300050); 4) ovarian cancer (OMIN *300050). Clones in this category include: fbr2_78k24; htes3_27d1.

Phosphoserine signature (phosphoglucomutases, phosphomannomutase): These proteins take part in the conversion of hexose phosphates. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following disease: Fanconi-Bickel Syndrome (OMIN #227810). Clones in this category include: fkd2_24b15.

NADH ubiquinone oxidoreductase: NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following disease: Brancio-oto-renal syndrome (OMIN *6601445). Clones in this category include: fkd2_3o17.

Transketolases: Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: Wernicke-Korsakoff Syndrome (OMIN *277730). Clones in this category include: tes3_17i17.

Fatty acid-CoA synthetases/ligases: These proteins contain AMP-binding domain signature(s), which is present in enzymes which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic,

causative, and/or related, etc...) with the following diseases: 1) Alport syndrome , mental retardation and elliptocytosis (OMIN *300157); 2) Adrenoleukodystrophy (OMIN *300100). Clones in this category include: tes3_35k17.

ADP/ATP or Adenine Nucleotide Translocators: These proteins contain mitochondrial energy transfer signature(s) and are most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) cardiomyopathy (OMIN *103220); 2) myopathy (OMIN *103220); 3) Progressive external ophthalmoplegia (OMIN *601227). Clones in this category include: tes3_35n12.

Carboxylesterases: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) hepatic carboxylesterase with detoxification of foreign compounds (OMIN *114835); 2) non-Hodgkin lymphoma (OMIN *114835); 3) B-cell chronic lymphocytic leukemia (OMIN *114835); 4) rheumatoid arthritis (OMIN *114835). Clones in this category include: tes3_35n9.

Heat shock proteins: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) 27 kd heat shock protein has been correlated with thermotolerance in response to environmental challenges and developmental transitions. (OMIN *6021295). Clones in this category include: utell_23e13.

Nucleic acid management

The genetic information is stored in the form of nucleic acids in all organisms. Two kinds of nucleic acids exist, DNA and RNA. Whereas the more stable DNA in most organisms constitutes the storage form of the genetic information, the labile RNA and in particular mRNA is an intermediate used for the temporal expression of specific genes.

In eukaryotes, DNA is usually a double stranded linear molecule consisting of two antiparallel strands and made up of a deoxyribose, a phosphorus backbone and the four bases A, C, G, and T. The DNA of some organisms has a ring structure. The structure of DNA was

unraveled years ago by Watson and Crick. DNA is directional molecule determined by the C-atoms of the sugar.

The most important processes dealing with nucleic acids are:

- replication (e.g. DNA polymerases, Telomerase)
- transcription (RNA polymerases)
- RNA processing (maturation - splicing and degradation)
- in addition, enzymes and proteins exist which require a nucleic acid (mostly RNA) in the active center to be functional (ribozymes - e.g. RNase, Ribosomal proteins)

The DNA of a cell is replicated in the S-phase of the cell cycle. Several enzymes carry out the task of doubling this nucleic acid. As all steps of the cell cycle, also the process of replication is tightly regulated. The enzyme DNA polymerase and several other proteins are involved in this process. Whereas many prokaryotes do have only one origin of replication (i.e., the starting point of the replication cycle), in eukaryotic DNAs (chromosomes) multiple such start points exist. The switch from the synthesis (S) phase to the subsequent G2 or M phases of the cell cycle are dependent on the completion of the replication. This makes clear, that a number of proteins are involved in the replication itself as well as in the control of the process. Since most eukaryotic chromosomes are linear structures, additional proteins and enzymes are necessary to make sure that the structure is maintained through successive generations. This includes those proteins necessary to build the three dimensional structure of chromosomes (e.g. histones) and the structural network of the nucleus and nucleolus (including the defined localization of transcriptionally active genes in the vicinity of nucleoli) but also such enzymes as telomerase which guarantees the integrity of the chromosomal ends.

The expression of genes is usually performed in two steps. First a messenger RNA (mRNA) is produced (transcribed) in one to many copies and second this mRNA is translated into the protein product. The regulation of transcription is discussed under the separate heading 'transcription factors', but also the classes 'signal transduction', 'development', 'cell cycle' and others are affected as the expression of certain genes determines the fate of a cell or organism.

The primary transcript (hnRNA - heterogeneous nuclear RNA) is a single stranded one-to-one copy of the gene as it is located on the chromosome. Before a protein can be translated, already during transcription the process of maturation is initiated. Firstly, a 5' cap structure is enzymatically and covalently added to the RNA, blocking the 5' end of the RNA.

Second, when the RNA polymerase has terminated polymerization, the enzyme poly A polymerase adds varying numbers of adenine residues to the 3' end of the transcript. This enzyme recognizes the sequence AAUAAA or AUUAAA (+ some minor variations), cuts the RNA 10 - 30 nucleotides downstream and adds the A residues. The size of the poly A sequence affects the stability of the RNA. Finally, in the process of splicing, the introns present on the genomic level and also present in the hnRNA are spliced out by a multi-protein complex consisting of several proteins and RNAs. The finally matured mRNA is exported to the cytoplasm where it is translated with help of the ribozymes.

The half life of RNA is usually much shorter than that of DNA. Usually, the mRNA is degraded shortly after synthesis, to guarantee a very defined window of expression of a given gene. This regulation is necessary to specifically maintain or change the set of proteins present at any time in a cell. Specific regions in the 3'UTR (untranslated region) determine the stability of the mRNA in the cytoplasm before it is degraded by RNases, enzymes consisting both of protein and RNA.

References: Watson and Crick (1953) *Nature* 171: 737-738.

Several categories of proteins are coded for by clones of the invention within the overall group of "Nucleic acid management" and include, among others, the following:

RNA helicases including DEAD/H box helicases: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. DEAD box proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by with the following disease processes and/or genes: 1) ataxia-telangiectasia gene: "A human gene (DDX10) encoding a putative DEAD-box RNA helicase at 11q22-q23" *Genomics* 33:199-206, 1996, Savitsky et al., (OMIN *601235); 2) hematopoietic tumors: "Cloning and expression of a murine cDNA homologous to the human RCK/P54, a lymphoma-linked chromosomal breakpoint 11q23", *Gene* 166:293-6, 1995, Seto et al. (OMIN *600326); 3) dermatomyositis: a) "The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation."

Arthritis Rheum. 38: 1389-1399, 1995, Seelig et al. (OMIN *603277); b) "Two forms of the major antigenic protein of the dermatomyositis-specific Mi-2 autoantigen." (Letter), *Arthritis Rheum.* 39: 1769-1771, 1996., Seelig et al. (OMIN *603277); c) "The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities", *Cell* 95: 279-289, 1998. Zhang et al. (OMIN *603277); 4) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200); 5) Mucopolysaccharidosis Type IVA (OMIN *253000); 6) Albinism I (OMIN *203100); 7) Wilms Tumor 1 (OMIN *194070); 8) Spinocerebellar Ataxia 7 (OMIN *164500). Clones in this category include: fbr2_23b10, fbr2_3cl8, fbr2_6o17, fbr2_82i24, and tes3_14h21.

Inorganic pyrophosphatase: Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity. Clones in this category include: fbr2_64a15.

DNA-damage –inducible protein (dinP) or Proteins induced by DNA-Damage: The dinB/P pathway is a second SOS-pathway in E.coli. Genes related to this seem to be involved in modulating DNA repair and mutagenesis. Clones in this category include: fbr2_72b18.

Proteins with myc-type, helix-loop-helix dimerization domain signature(s). This helix-loop-helix domain mediates protein dimerization has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, these proteins could be novel DNA-binding proteins. Clones in this category include: fbr2_72i12.

Cytosolic ribosomal proteins L36: L36 seems to be part of the eukaryotic ribosomal peptidyl transferase center and can find application in modulation of ribosome assembly, maintenance and activity. Clones in this category include: fkd2_3b2.

Ribonuclease H: Ribonuclease H proteins are RNA modifying proteins and have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Adenomatous Polyposis of the Colon (OMIN

*175100); 2) Retinoblastoma (OMIN *180200) ; and 3) Von Hippel-Lindau Syndrome (OMIN *193300). Clones in this category include: phtes3_15j3.

Signal transduction

Cells in higher order organisms need to continuously communicate with its environment especially with other cells of the same organism in order to maintain the function and specialization of the whole system these cells are part of. This important task of communication is performed with help of cell-surface receptors which receive and transmit signals from outside into the cell.

G-proteins

The largest known family of cell-surface receptors is that of the G-protein-coupled receptors, which mediate the transmission of diverse stimuli such as neurotransmitters, glycopeptides, hormones, peptides, odorant molecules, and photons. The functional unit of these receptors is composed of the receptor molecule itself (GPCR) which is anchored in the cytoplasmic membrane with seven membrane spanning domains, the heterotrimeric G-protein which is composed of α and $\beta\gamma$ -subunits ($G\alpha$ and $G\beta\gamma$), and the effectors that interact with $G\alpha$ and / or $G\beta\gamma$. In particular, the dissociated $G\alpha$ and $G\beta\gamma$ can regulate the activities of a number of effector molecules such as adenylate cyclases, phospholipase C isoforms, ion channels, and tyrosine kinases, resulting in a variety of cellular functions. The process of signal transduction must be tightly regulated and reversible in order to avoid overstimulation, to achieve signal termination, and render the receptor responsive to subsequent stimuli [Iacovelly L. et al., (1999) *FASEB J.* 13, 1-8, Hamm, H.E. (1998) *J. Biol. Chem.* 273, 669-672].

G-proteins are GTPases that, upon binding of GTP change their conformation which in return unmask structural motives, in particular the so called effector loop, which can mediate the interactions to target proteins, or effectors, for the GTPases. This ability enables the GTPases to cycle between active, GTP-bound and inactive, GDP bound conformations and in the process to function as molecular traffic lights in a multitude of signal transduction pathways. The most important of these signal transduction pathways that are regulated with help of G-proteins are that of the phospholipase C / protein kinase C and that of the adenylate cyclase / protein kinase A.

The cycling of GTPases is tightly regulated by three main classes of proteins: The exchange of hydrolyzed GDP for a fresh GTP is facilitated by guanosine nucleotide exchange factors (GEFs), the hydrolysis of GTP to GDP is sped up by GTPase-activating proteins (GAPs), and the dissociation of GDP from the GTPases is inhibited by GDP dissociation inhibitors (GDIs) [Tapon and Hall (1997) *Curr. Opin. Cell. Biol.* **9**, 86-92, Van Aelst and D-Souza-Schorey (1997) *Genes Dev.* **11**, 2295-2322].

SOC-family

A conserved motif that was originally identified in proteins that negatively regulate the signaling action of cytokines was termed SOCS box, the Suppressor Of Cytokine Signaling. Based on homology, five distinct structural protein classes have been identified since that carry this motif. The function of most of these proteins is presently not known. Common to the proteins is only the SOCS box which is located near the C-terminus of the respective peptides. Recently, the SOCS box has been demonstrated to induce binding of proteins to elongins B and C which could target the proteins (and bound substrates) to the proteasomal protein degradation pathway (Kamura, T. *et al.* (1998) *Genes Dev.* **12**, 3872-3881; Zhang, J.-G. *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2071-2076).

The class where the SOCS box was originally described contains several members (SOCS-1-SOCS-7 and CIS). In addition to the SOCS box, these proteins also contain a SH2 (Src-homology 2) domain and a variable N-terminus. These SOCS proteins appear to form part of a classical negative feedback loop that regulates cytokine signal transduction. Upon cytokine stimulation, expression of SOCS proteins is rapidly induced and the proteins inhibit further cytokine action. The mode of action of the SOCS proteins is variable. While SOCS-1 binds and inhibits the JAK (Janus kinases) family of cytoplasmic protein kinases [Narahzaki M. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13130-13134, Nicholson, S.E. *et al.* (1999) *EMBO J.* **18**, 375-385], CIS appears to act by competing with signaling molecules such as the STATs (Transducers and Activators of Transcription) family for binding to phosphorylated receptor cytoplasmic domains [Yoshimura, A. *et al.* (1995) *EMBO J.* **14**, 2816-2826; Matsumoto, A. *et al.* (1997) *Blood* **89**, 3148-3154].

A second class of SOCS box protein contains additionally WD-40 repeats which were initially identified in the mouse WSB-1 and -2 proteins. The functions of WD-40 proteins are not completely understood but seem to be rather divergent. In Cdc4p the WD-40 repeats probably are necessary for binding the substrate for Cdc34p [Mathias, N. *et al.* (1999) *Mol.*

Cell Biol. **19**, 1759-1767]. Cdc4p is a component of a ubiquitin ligase that tethers the ubiquitin-conjugating enzyme Cdc34p to its substrates. The posttranslational modification of a protein by ubiquitin usually results in rapid degradation of the ubiquitinated protein by the proteasome. The transfer of ubiquitin to substrate is a multistep process where WD-40 repeats might play an important function.

Other WD-40 containing proteins (e.g. the retino blastoma binding protein RbAp48) have been shown to bind metal ions (Zinc) and that this metal binding might mediate and/or regulate protein-protein interactions which are functionally important in chromatin metabolism [Kenzior, A.L. and Folk, W.R. (1998) *FEBS Lett.* **440**, 425-429]. These proteins are involved in the RAS-cAMP pathway that regulates cellular growth [Ach R.A. *et al.* (1997) *Plant Cell* **9**, 1595-1606].

The SPRY domain has been identified in pyrin or marenostrin, a protein which is mutated in patients with Mediterranean fever and which is similar to the butyrophilin family. While butyrophilins seem to be involved in the lactation process in mammals, the function pyrin is unknown. Three proteins (SSB-1 to -3) have been identified to contain both SPRY and SOCS box motifs. The function of these proteins is also not known.

Ankyrin repeat containing proteins share a 33-residue repeating motif, an L-shaped structure with protruding β -hairpin tips which mediate specific macromolecular interactions with cytoskeletal, membrane, and regulatory proteins. These proteins play fundamental roles in diverse biological activities including growth and development, intracellular protein trafficking, the establishment and maintenance of cellular polarity, cell adhesion signal transduction, and mRNA transcription. Three proteins that contain ankyrin repeats (ASB-1 to -3) have been identified to contain a C-terminal SOCS box additionally to the ankyrin repeats. The function of these proteins or the individual domains remains to be discovered [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

A few small GTPases (RAR and RAR like) do also contain a SOCS box. GTPases are involved in signal transduction during cellular communication. The function of the SOCS box in this type of proteins is currently unclear [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

Ca²⁺ as second messenger

The bivalent cation Ca²⁺ is, besides cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very

low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+} /calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+} . In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

cAMP

The cyclic AMP is produced by the enzyme adenylate cyclase in response to extracellular signals. Certain G-proteins stimulate the activity of adenylate cyclase which converts ATP to cAMP and PPi. Two molecules of cAMP bind to each of two regulatory subunits of cAMP dependent protein kinase which in turn dissociate from the two catalytic subunits of the heterotetramer R_2C_2 . Upon release of the C-subunits, they become active and phosphorylate substrate proteins at Ser and Thr residues. The process leading from binding of extracellular molecules to their receptors, the transmission of the stimuli into the cell, the activation of adenylate cyclase and the subsequent activation of cAMP dependent protein kinase is one of two major signal transduction pathways in eukaryotic cells. Since the phosphorylation of proteins is a posttranslational modification of proteins, the kinases are described in the class "signal transduction."

SARA

Members of the transforming growth factor β (TGF β) superfamily signal through a family of cell-surface transmembrane serine/threonine kinases, known as type I and type II receptors (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Ligand induces formation of heteromeric complexes of these receptors, and signaling is initiated when receptor I is phosphorylated and activated by the constitutively active kinase of receptor II (Wrana et al., 1994). The activated type I receptor kinase then propagates the signal to a family of intracellular signaling mediators known as Smads (contraction of the C.elegans Sma and Drosophila Mad genes which were the first identified members of this class of signaling effectors).

Three classes of Smads with distinct functions have been defined: the receptor-regulated Smads, which include Smad1, 2, 3, 5, and 8; the common mediator Smad, Smad4; and the antagonistic Smads, which include Smad6 and 7 (Heldin et al., 1997; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Receptor-regulated Smads (R-Smads) act as direct substrates of specific type I receptors, and the proteins are phosphorylated on the last two serines at the carboxyl terminus within a highly conserved SSXS motif (Macías-Silva et al., 1996 ; Abdollah et al., 1997 ; Kretzschmar et al., 1997 ; Liu et al., 1997b ; Souhelnytskyi et al., 1997). Regulation of R-Smads by the receptor kinase provides an important level of specificity in this system. Thus, Smad2 and Smad3 are substrates of TGF β or activin receptors and mediate signaling by these ligands (Macías-Silva et al., 1996 ; Liu et al., 1997b ; Nakao et al., 1997), whereas Smad1, 5, and 8 are targets of BMP receptors and propagate BMP signals (Hoodless et al., 1996 ; Chen et al., 1997b ; Kretzschmar et al., 1997 ; Nishimura et al., 1998). Once phosphorylated, R-Smads associate with the common Smad, Smad4 (Lagna et al., 1996 ; Zhang et al., 1997), and mediate nuclear translocation of the heteromeric complex. In the nucleus, Smad complexes then activate specific genes through cooperative interactions with DNA and other DNA-binding proteins such as FAST1, FAST2, and Fos/Jun (Chen et al., 1996 , Chen et al., 1997a ; Liu et al., 1997a ; Labbé et al., 1998 ; Zhang et al., 1998 ; Zhou et al., 1998). In contrast to R-Smads and Smad4, the antagonistic Smads, Smad6 and 7, appear to function by blocking ligand-dependent signaling (reviewed in Heldin et al., 1997).

Phosphorylation of R-Smads by the type I receptor is essential for activating the TGF β signaling pathway (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). However, little is known of how Smad interaction with receptors is controlled. A novel Smad2/Smad3 interacting protein has been described (Tsukazaki T. et al., 1998) that contains a double zinc finger, or FYVE domain, and which has been called SARA (Smad ancor for receptor activation). The SARA motif recruits Smad2 into distinct subcellular domains and co-localizes and interacts with TGF β receptors. TGF β signaling induces dissociation of Smad2 from SARA with concomitant formation of Smad2/Smad4 complexes and nuclear translocation. Moreover, deletion of the FYVE domain in SARA causes mislocalization of Smad2 and inhibits TGF β -dependent transcriptional responses. Thus, SARA defines a component of TGF β signaling that functions to recruit Smad2 to the receptor by controlling the subcellular localization of Smad.

References: Abdollah et al. (1997) *J. Biol. Chem.* 272, 27678-27685; Attisano et al. (1998) *Curr. Opin. Cell Biol.* 10, 188-194; Chen et al. (1996) *Nature* 383, 691-696; Chen et al. (1997a) *Nature* 389, 85-89; Chen et al. (1997b) *Proc. Natl. Acad. Sci. USA* 94, 12938-12943; Heldin et al. (1997) *Nature* 390, 465-471; Hoodless et al. (1996) *Cell* 85, 489-500; Kretschmar et al. (1998) *Curr. Opin. Genet. Dev.* 8, 103-111; Kretschmar et al. (1997) *Genes Dev.* 11, 984-995; Labbé et al. (1998) *Mol. Cell* 2, 109-120; Lagna et al. (1996) *Nature* 383, 832-836; Liu et al. (1997a) *Genes Dev.* 11, 3157-3167; Liu et al. (1997b) *Proc. Natl. Acad. Sci. USA* 94, 10669-10764; Macías-Silva et al. (1996) *Cell* 87, 1215-1224; Nakao et al. (1997) *EMBO J.* 16, 5353-5362; Nishimura et al. (1998) *J. Biol. Chem.* 273, 1872-1879; Souchelnytskyi et al. (1997) *J. Biol. Chem.* 272, 28107-28115; Tsukazaki et al. (1998) *Cell* 95, 779-791; Wrana et al. (1994) *Nature* 370, 341-347; Zhang et al. (1997) *Curr. Biol.* 7, 270-276; Zhang et al. (1998) *Nature* 394, 909-913; Zhou et al. (1998) *Mol. Cell* 2, 121-127.

Calcium

The bivalent cation Ca^{2+} is, along with cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+} /calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+} . In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

Rab proteins

In eukaryotic cells the compartmentalization of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins and other molecules. Trafficking between organelles within the secretory pathway occurs as

vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated α -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes *in vitro*, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle,

most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown in vitro to interact with -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca^{2+} -binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca^{2+} influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits

that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn²⁺-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991). *Mol. Cell. Biol.* 11, 872-885; Echard et al. (1998). *Science*. 279, 580-585; Geppert et al. (1998). *Annu. Rev. Neurosci.* 21, 75-95; Guo et al. (1999). *EMBO J.* 18, 1071-1080; Kato et al. (1996). *J. Biol. Chem.* 271, 31775-31778; Novick et al. (1997). *Curr. Opin. Cell Biol.* 9, 496-504; Peterson et al. (1999). *Curr. Biol.* 9, 159-162; Poirier et al. (1998). *Nat. Struct. Biol.* 5, 765-769; Vitale et al. (1998). *EMBO J.* 17, 1941-1951; Wang et al. (1997). *Nature*. 388, 593-598; Yang et al. (1999). *J. Biol. Chem.* 274, 5649-5653.

Kinases

Reversible posttranslational modifications of proteins are major means of regulating cellular activities. Among the various modifications that are carried out by the cells, the addition of phosphoryl groups to Ser/Thr or Tyr residues is the most important and widely used. The phosphorylation of proteins is accomplished by protein kinases, while the reverse reaction, the removal of phosphoryl groups, is carried out by phosphatases. Kinases / Phosphatases regulate key positions e.g. in the processes of cell proliferation, differentiation and communication/signaling. These processes must be tightly regulated in order to maintain a steady state level of cellular fate. Mis-regulation of kinase activities (or that of

phosphatases) is made responsible for a multitude of disease processes such as oncogenesis, inflammatory processes, arteriosclerosis, and psoriasis.

Protein kinases constitute the largest protein family that is currently known. Several hundred kinases have been identified already. Classically, kinases are subdivided into two classes based on the amino acid residues in their substrates that are phosphorylated by the particular enzymes. The kinases specifically add phosphoryl groups from adenosine triphosphate (ATP) or, less frequently, guanosine triphosphate (GTP), either to serine and/or threonine or to tyrosine residues of substrate proteins. An estimated 1,000 to 10,000 proteins present in a typical mammalian cell are believed to be regulated also by the action of protein kinases.

Protein kinases are frequently integral parts of signaling cascades that transmit extracellular stimuli (e.g. hormones, neurotransmitters, growth- or differentiation factors) into the cell and result in various responses by the cells. The kinases play key roles in these cascades as they constitute a sort of 'molecular switches' turning on or off the activities of other enzymes and proteins, e.g. metabolic, regulatory, channels and pumps, receptors, cytoskeletal, transcription factors.

The regulation of kinase activities is accomplished by various means:

The best characterized example for the regulation via regulatory subunits is the cAMP-dependent protein kinase (PKA) which is also a prototype for second messenger activated protein kinases. This enzyme consists of a heterotetramer of two catalytic (C) and two regulatory (R) subunits. Upon binding of two molecules of second messenger (cAMP) in each R subunit, the catalytic subunits are released and active. Both of the catalytic and the regulatory subunits several isoforms exist. The combination of catalytic and regulatory subunits determines the localization of the holoenzyme and also the substrate spectrum that is available for phosphorylation. The consensus pattern necessary to be present in the substrate for PKA action is RRXS/T where X can be any amino acid.

The casein kinase II comprises another examples for holoenzymes that consist of catalytic and regulatory subunits. Other kinases that are activated by second messengers are cGMP-dependent protein kinase and Protein kinase C (PKC) which is activated by diacylglycerol, which in turn is produced by phospholipases by cleavage of phosphatidylcholine.

Receptor kinases usually consists of an extracellular domain which can bind effector molecules (e.g. growth factors and hormones) and transfer the stimulus to the intracellular domain of these proteins which usually is a protein tyrosine kinase. Other tyrosine kinases lack an extracellular domain but are associated with receptors which transfer the signal after effector binding by activating the associated protein kinase enzyme (e.g. Src kinase family; Src, Blk, Fgr, Fyn, Lck Lyn, Yes and Janus kinase family; Jak1-3, Tyk2).

Dysfunction of kinases, e.g. caused by non-functioning regulation, can be the cause of inflammatory diseases and uncontrolled proliferation. v-Src which is a truncated version of the C-Src protooncogene tyrosine kinase is a classical example for this process as v-Src does not contain the regulatory domain of the cellular gene and is thus constitutively active.

Several categories of proteins are coded for by clones of the invention within the overall group of "Signal transduction" and include, among others, the following:

Neurocalcin (Recoverin): Neurocalcin is a Ca^{2+} -binding protein with three putative Ca^{2+} -binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca^{2+} dependent activation of guanylate cyclase.. These proteins can find application in modulating/blocking the guanylate cyclase-pathway. Diseases associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) autosomal dominant cone dystrophy (OMIN *600364); 2) cone dystrophy 3 (OMIN *600364); 3) cancer associated retinopathy (OMIN *179618). Clones in this category include: fbr2_23b21.

Proteins with a WW Domain: Proteins that contain a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore these proteins should be involved in intracellular signal transduction. Diseases associated (as

potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200). Clones in this category include: fbr2_23n16.

Protein substrates for cAMP-dependent protein kinase: Acting as a choride channel or chloride channel inhibitor these proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Cystic Fibrosis (OMIN #219700). Clones in this category include fbr2_82i17.

Sphingosine kinase: Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellular, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependent on SPP. Extracellular, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1. These proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Gaucher Disease, Type I (OMIN *230800). Clones in this category include fbr2_82m6.

Vanilloid Receptors: VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. Related can find application as a target for the development of new nociception-modulating drugs. Clones in this category include tes3_20k2.

RCC1 (Regulator of chromosome condensation): RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator. These proteins can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat. OMIN also reports that RCC1 has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with retinitis pigmentosa (OMIN *312610). Clones in this category include tes3_21d4.

Ras inhibitor proteins: Ras is a signal transducing molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show

intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. Ras inhibitor proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with many disease processes as reported by OMIN including: 1) Tumors of the lung, breast, brain, pituitary, pancreas, bone, skin, bladder, kidney, ovary, prostate and lymphocyte, Melanoma (OMIN *600160); 2) X-linked non-specific mental retardation (OMIN *300104); 3) adenomatous polyposis of the colon (OMIN *175100); 4) Beckwith-Wiedemann Syndrome (#130650); and 5) Major affective disorder 1 (OMIN *125480). Clones in this category include ute1_22g21.

Mammalian proteins cornicon involving the EGF-receptor: Cornicon proteins are part of a signal transduction pathway involving the EGF-receptor. The EGF-receptor has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Leprechaunism (OMIN #246200); 3) Hemophilia B (OMIN *306900); 4) Ectodermal dysplasia 1; 5) Kartagener syndrome (OMIN *244400) and 6) Glioma of the brain (OMIN *137800).). Clones in this category include ute1_22e12.

Transmembrane proteins

Membrane region prediction was effected using the ALOM2 software (Klein et al., 1985; version 2 by K. Nakai). Similar to many other methods, the Kyte & Doolittle (1982) amino acid hydrophobicity scale is used in ALOM2 as the primary variable for classifying sequences in terms of their localization. High prediction accuracy is achieved through the system of intelligent decision rules and the utilization of a carefully selected training data set. The method also generates reliability estimates which makes it possible to distinguish between membrane-spanning proteins (I, intrinsic) and globular proteins with regions of high hydrophobicity buried in the core.

For a protein of length L , the block of length l with maximum hydrophobicity is found:

$$\max H = \max(1/l) \sum_{i=k}^{k+l-1} H_i$$

$k=1, \dots, L-l+1$

where H_i represents the hydrophobicity of an individual residue.

Let $P(I/\max H)$ and $P(E/\max H)$ be the conditional probabilities that a protein is integral or peripheral, respectively, given its value of maximal hydrophobicity $\max H$, and let $P(I)$ and $P(E)$ be the prior probabilities of intrinsic and extrinsic membrane proteins estimated from the training set. Then a sequence is assigned to E if

$$P(E/\max H) > P(I/\max H)$$

or, after applying the Bayes rule,

$$P(E)P(\max H/E) > P(I)P(\max H/I),$$

where the conditional probabilities $P(\max H/E)$ and $P(\max H/I)$ can be determined based on the estimates of probability distributions of $\max H$ in both groups.

Discriminant analysis allows to simplify this task by calculating the odds $P(E/\max H):P(I/\max H)$ as e^b , where b is the left-hand side of a linear or quadratic inequality. For example, for the window of length 17, the protein is allocated to the peripheral category E based on the empirically derived quadratic inequality:

$$1.05(\max H)^2 + 12.30\max H + 17.49 > 0,$$

whereas the optimal inequality for assigning membrane proteins (category I) is linear:

$$-9.02\max H + 14.27 > 0$$

The odds parameter can be made more or less stringent. For example, one can require odds at least 1:10 for a protein to be classified as integral. This leads to higher selectivity but less sensitivity.

The boundaries of membrane-spanning regions in putative membrane proteins are detected by means of an iterative procedure whereby the most hydrophobic region corresponding to the value $\max H$ is considered to be membrane and removed from the sequence. The classification procedure is then repeated again for the remaining sequence, and, if such a protein is again classified as integral, the next most hydrophobic region is considered.

Reference: Klein, P., Kanehisa, M., DeLisi, C. (1985) The detection and classification of membrane-spanning proteins. *Biochem Biophys Acta* **815**: 468-476

Transcription factors

Purified eukaryotic RNA polymerase II is unable to initiate promoter-specific transcription. A family of factors that collectively confer RNAPII promoter specificity is known as the general transcription factors (GTFs). They include the TATA-binding Protein (TBP) TFIIB, TFIIE, TFIIIF and TFI IH. These factors are conserved among all eukaryotes.

RNAPII complexes containing the entire set of GTFs or a subset of GTFs together with other proteins have been isolated from mammalian and yeast cells. Although purified RNAPII and GTFs are sufficient for promoter-specific initiation, this system fails to respond to activators. This is mediated by a further complex termed mediator complex which associates with the carboxy-terminal heptapeptide domain (CTD) of the largest subunit of RNAPII.

Purification of human RNAPII complexes resulted in two distinct forms of human RNAPII after analysis of functional properties. One complex contained chromatin remodeling activities but was devoid of GTFs. The other complex did not contain factors that modify chromatin but contained a subset of SRB/mediator subunits and GTFs and other polypeptides that mediate transcriptional activation, a scenario similar to that reported for yeast.

A complex designated NAT (~20 SU) for negative regulator of transcription contains RNAPII, Cdk8, homologs of the yeast mediator complex as well as Rgr1 and Srb10/11 known as negative regulators of transcription.

A complex with striking similar structural and functional properties to NAT has been identified designated SMCC (~15 SU) (SRB/mediator coactivator complex), that can also mediate transcriptional activation.

The SMCC complex includes all reported NAT subunits including subunits of the TRAP complex. TRAP is a coactivator complex isolated on the basis of its interaction with the thyroid hormone receptor. Another coactivator complex DRIP, isolated on the basis of its

ability to interact with the vitamin D3 receptor, contains novel subunits as well as subunits of NAT/SMCC and TRAP complexes.

The effects of each of these coactivator complexes is dependent on the TFIID complex. It is not known if the TAF subunits of TFIID are required. It is likely that new coactivator complexes will be uncovered containing both novel and previously defined components.

Beside the huge amount of transcription factors which can be part of the RNAIIP holoenzyme or the coactivator complexes there is an even larger quantity of specific transcription factors binding to promoter elements within the DNA sequences of a given gene leading to activation or repression of transcription. A broad range of cellular responses like differentiation, proliferation, cell death and others are elicited through activating or repressing the transcription of target genes.

There are at least five superclasses of transcription factors:

1. Superclass contains members with characteristic basic domains:

Members are:

Leucine zipper factors, where the basic domain is followed by a leucine zipper of repeated leucine residues at every seventh position. The zipper mediates protein dimerization as a prerequisite for DNA-binding.

Helix-loop-helix factors (bHLH) contain a DNA-binding basic region followed by a motif of two potential amphipathic alpha-helices connected by a loop of variable length also mediating dimerization.

Factors with a combination of Helix-loop-helix and leucine zipper.

Further members of this superclass are NF-1, RF-X, and bHSH like proteins.

2. Superclass comprises factors containing zinc-coordinating DNA-binding domains.

Members are:

Proteins with Cys4 zinc finger of nuclear receptor type, where two such motifs differing in size, composition and function are present in each receptor molecule. Each finger comprises 4 cysteine residues coordinating one zinc ion. The second half including the second cysteine pair has alpha-helix conformation and the helix of the first finger binds to the DNA through the major groove. The sequence between the first two cysteines of the second finger mediates dimerization upon DNA-binding. This class includes the steroid hormone receptors and the thyroid hormone receptor-like factors. Other diverse cys4 zinc fingers have a motif of GATA-type.

Proteins with Cys2His2 zinc finger domain(s). Each finger comprises 2 cysteine and 2 histidine residues coordinating one zinc ion, and in some cases one histidine is replaced by another cysteine. The zinc ion is essential for DNA-binding.

Proteins with Cys6 cysteine-zinc cluster(s). Six cysteine residues coordinate two zinc ions, i. e. two of the thiol groups are coordinating two zinc ions each. Present in many fungal regulators.

Zinc fingers of alternating composition.

3. Superclass contains factors of helix-turn-helix type.

Members are:

Proteins with homeo domains. Homeo domains are three consecutive alpha-helix structures. Helix 3 contacts mainly the major groove of the DNA, some contacts at the minor groove are observed as well. Helix 2 and 3 resemble the helix-turn-helix structure of prokaryotic regulators.

Proteins with Paired box domain(s). This is a DNA-binding domain of approximately 130 amino acid residues. Its N-terminal half is basic, its C-terminal half is highly charged in general. It probably comprises 3 alpha-helices.

Proteins with Fork head / winged helix domain(s). This domain was identified by homology between HNF-3A and flk. The domain comprises approx. 110 AA. Analysis of the crystal structure has revealed a compact structure of three alpha-helices, the third alpha-helix

being exposed towards the major groove of the DNA. The domain also exerts minor groove contacts. Upon binding to DNA, it induces a bend of 13 degree.

Heat shock factors

Proteins with Tryptophan clusters. The tryptophan clusters comprise several tryptophan residues with a spacing of 12-21 amino acid residues; the subclass of myb-type DNA-binding domains typically exhibit a spacing of 19-21 amino acid residues.

Proteins with TEA domain(s). The TEA domain has been identified as a region which is conserved among the transcription factors TEF-I, TECl and abaA. This domain in TEF-I has been shown to interact with DNA, although two additional regions may also contribute to DNA-binding. It is predicted to fold into three alpha-helices, with a randomly coiled region of 16-18 amino acid residues between helices 1 and 2, and a short stretch between helices 2 and 3 of 3-8 residues.

4. Superclass contains beta-Scaffold Factors with Minor Groove Contacts

Members are:

Proteins with RHR (Rel homology) region.

The structure of the Rel-type DBD exhibits a bipartite subdomain structure, each subdomain comprising a beta-barrel with five loops that form an extensive contact surface to the major groove of the DNA. Particularly, the first loop of the N-terminal subdomain (the highly conserved recognition loop) performs contacts with the recognition element on the DNA, but other loops are involved. The fact that the main DNA-contacts are made through loops has been suggested to provide a high degree of flexibility in binding to a range of different target sequences. Augmenting interactions are achieved by two alpha-helices within the N-terminal Part that form strong minor groove contacts to the A/T-rich center of the B-element. In p65, the sequence between both alpha-helices is much shorter and even helix 2 is truncated. The second, C-terminal domain is necessary mainly for protein dimerization.

p53 proteins

MADS (MCM1-agamous-deficiens-SRF) box proteins. Proteins of this class comprise a region of homology. The DNA-binding domain also comprises the dimerization capability. In the DNA-bound dimer (shown for SRF), two antiparallel amphipathic alpha-helices (alpha-I), form a coiled coil and are oriented approximately parallel on the minor groove. These helices make minor and major groove contacts, the N-terminal extensions form minor groove contacts. The bound DNA is bent and wrapped around the protein. It exhibits a compressed minor groove in the center and widened minor groove in the flanks.

Beta-Barrel alpha-helix transcription factors.

TATA-binding proteins

HMG proteins

Proteins of this class comprise a region of homology with the chromosomal non-histone HMG proteins such as HMG1. This region comprises the DNA-binding domain which in some instances such as HMG1 mediates sequence-unspecific, in other cases such as LEF-1 sequence-specific binding to DNA. This domain exhibits a typical L-shaped conformation made up of 3 alpha-helices and an extended N-terminal extension of the first helix. The latter together with helix 1, which contains a kink, form the long arm of the L, whereas helices 1 and 2 form the short arm. Binding to the minor groove induces a sharp bending of the DNA by more than 90 degree, away from the bound protein. The overall topology of the DNA-protein complexes resembles somewhat that of the TBP-TATA box complex.

Heteromeric CCAAT factors

Proteins with Grainyhead domain(s)

Cold-shock domain factors. Cold-shock domain proteins are characterized by a highly conserved region first found in prokaryotic cold-shock proteins. This domain is a single-stranded nucleic acid-binding structure interacting with DNA or RNA. It consists of an antiparallel five-stranded beta-barrel, the strands of which are connected by turns and loops. Within this structure, a three-stranded beta-strand contains a conserved RNA-binding motif, RNPI. Not all CSD proteins are transcription factors. Those which specifically bind to a

certain sequence are termed Y-box proteins. Proteins of this class were previously called protamine-like domain proteins because of having a highly positively charged domain with interspersed proline residues.

Proteins with Runt homology domain

The members of this transcription factor class have been identified on the basis of their homology to a defined region within the *Drosophila* protein Runt. The runt domain is part of the DNA-binding domain of these factors. It consists mainly of beta-strands, does not contain alpha-helical regions and seems to be most similar to the palm domain found in DNA polymerase beta (rat).

5. Superclass contains other transcription factors like Copper fist proteins, HMGI(Y), STAT, Pocket domain proteins and Ap2/EREBP-related factors.

The classification of transcription factors originates from TRANSFAC database:

<http://transfac.gbf.de/TRANSFAC/>

Reference: Heinemeyer

Several categories of proteins are coded for by clones of the invention within the overall group of "Transcription Factors" and include, among others, the following:

Dcoh: Dcoh is a bifunctional protein, complexed with bipterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the bipterin cofactor of phenylalanine hydroxylase. The Dcoh protein has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) hyperphenylalanemia (OMIN 126090, #264070). Clones in this category include fkd2_46k12.

Signal transducing proteins: Beta-transducin subunits of G-proteins contain WD-40 repeats. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription. These proteins have been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) essential hypertension (OMIN *139130). Clones in this category include ute1_1i2.

* * *

The invention, therefore, specifically contemplates the following assemblages of materials, which track the above-identified fourteen functional groupings, that are useful in practicing the profiling aspects of the invention. One type of assemblage is nucleic acid-based and can include the following groupings of sequences and their derivatives: all sequences; human fetal brain sequences; brain derived sequences; human fetal kidney library sequences; kidney derived sequences; human mammary carcinoma library sequences; mammary carcinoma derived sequences; human testis library sequences; testes derived sequences; cell cycle genes; cell structure and motility genes; differentiation and development genes; intracellular transport and trafficking genes; metabolism genes; nucleic acid management genes; signal transduction genes; transmembrane protein genes; and transcription factor genes. Other assemblages contain proteins or their corresponding antibodies or antibody fragments, divided along the same groupings.

Database Applications

Because they are human genes and gene products, the inventive molecules are useful as members of a database. Such a database may be used, for example, in drug discovery and rationale drug design or in testing the novelty and non-obviousness of newly sequenced materials. In addition, they are particularly suited in designing variants for the profiling (and other) applications described herein. Hence, the following discussion of electronic embodiments applies equally to such variants, which, naturally, will be generated and stored using a computer using known methodologies.

Accordingly, one aspect of the invention contemplates a database of at least one of the inventive sequences stored on computer readable media. Again, the individual sequences may be grouped with regard to the individual functional and structural groups mentioned above. While the individual sequences of a database may exist in printed form, they are preferably in electronic form, as in an ascii or a text file. They may also exist as word processing files or they may be stored in database applications like DB2, Sybase, Oracle, GCG and GenBank. One skilled in the art will understand the range of applications suitable for using and storing the electronic embodiments of the invention.

“Computer readable media” refers to any medium which can be read and accessed by a computer. These include: magnetic storage media, like floppy discs, hard drives and magnetic tape; optical storage media, like CD-ROM; electrical storage media, like RAM

and ROM; and hybrids of these categories, like magnetic/optical storage media. One skilled in the art will readily understand the scope of computer readable media and how to implement them.

Biological Activities and Assays for Implementing Therapeutic and Diagnostic Applications

This section provides assays for biological activity that are useful in characterizing and quantifying the biological activity of the inventive molecules and their derivatives, which is relevant to the pharmacological effects of the inventive molecules. As used in this section, it will be understood that "protein" may also refer to the inventive antibodies (including fragments).

Cytokine and Cell Proliferation/Differentiation Activity

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M + (preB M +), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin gamma, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6-Nordan, R. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9-Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

Immune Stimulating or Suppressing Activity

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to modify immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the

tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., *Science* 257:789-792 (1992) and Turka et al., *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function in vivo on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-vital immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient.

The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection in vivo.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and beta 2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Herrmann et al., *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann et al., *J. Immunol.* 128:1968-1974, 1982; Handa et al., *J. Immunol.* 135:1564-1572, 1985; Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988; Herrmann et al., *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann et al., *J. Immunol.* 128:1968-1974, 1982; Handa et al., *J. Immunol.* 135:1564-1572, 1985; Takai et al., *J. Immunol.* 137:3494-3500, 1986; Bowman et al., *J. Virology* 61:1992-1998; Takai et al., *J. Immunol.* 140:508-512, 1988; Bertagnolli et al., *Cellular Immunology* 133:327-341, 1991; Brown et al., *J. Immunol.* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; and Assays for B cell function: *In vitro* antibody production, Mond, J. J. and Brunswick, M. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988; Bertagnolli et al., *J. Immunol.* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., *J. Immunol.* 134:536-544, 1995; Inaba et al., *Journal of*

Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

Tissue Growth Activity

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the

treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and

cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

Activin/Inhibin Activity

A protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle

stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin alpha family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin- beta group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., *Endocrinology* 91:562-572, 1972; Ling et al., *Nature* 321:779-782, 1986; Vale et al., *Nature* 321:776-779, 1986; Mason et al., *Nature* 318:659-663, 1985; Forage et al., *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986.

Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of

cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such

receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of

cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in

a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Particular Applications for Certain Clones

The following sets out a non-exclusive list of applications for certain embodiments of the invention. In the interest of economy, applications relevant to multiple embodiments are not duplicated in this list. Other embodiments described in below have similar characteristics, as described therein. The artisan is directed, therefore, to this section for similar descriptions of the functions of other embodiment.

Testes

htes3_15c24: The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

htes3_15i5: The new protein can find application in modulating the structure of the human spermatozoa radia spoke head and modulation of sperm motility in men.

htes3_15k11: The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

htes3_17n12: The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

htes3_20k2: The new protein can find application as a target for the development of new nociception-modulating drugs.

htes3_20m18: The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

htes3_20d4: The new protein can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

htes3_21j15: NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C₂H₂ Zinc fingers and is a new putativ transcription factor. The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

htes3_26g22: The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP. The new protein can find application in modulation of gene transcription.

htes3_21i16: The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

htes3_27d1: The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

htes3_2m18: The novel protein can find application as multifunctional nuclease / exoribonuclease.

htes3_35b4: The new protein can find application in modulation of the mitotic spindle.

htes3_35b5: The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

htes3_35e21: Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells. This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

htes3_35k16: Therefore it is a new fatty acid-CoA synthetases/ligase with unknown substrate. The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

htes3_35n12: The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

htes3_35n9: The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

htes3_35p22: The novel protein is closely related to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control. The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

htes3_4h6: The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

htes3_72k15: FGD1-related F-actin-binding protein (Frabin/FGD1) is a novel F-actin-binding protein. The gene locus *fgd1* seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an *esin* yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

htes3_72p16: As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (PrA), proteinase B (PrB), and alkaline phosphatase (ALP). The new protein can find application in modulation the sorting of proteins into different compartments.

htes3_7b22: The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

htes3_7j3: The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too. The new protein can find application in modulating/blocking the cell cycle.

htes3_7p9: The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle. The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819) complex. The new protein can find application in modulation of viral infections and tumour events.

htes3_8m10: The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA. The new protein can find application in modulation of mRNA translation and processing/stability.

Kidney

hfkd2_24b15: The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

hfkd2_24n20: The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

hfkd2_3o17: The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

hfkd2_46j20: The new protein can find application in modulating the homoprotocatechuate degradative pathway and as a enzyme for biotechnologic production processes.

hfkd2_46k19: The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

hfkd2_46m4: SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

hfkd2_46k14: rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport. The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

Uterus Associated:

hutel_18i19: The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH2-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

hutel_18l1: The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome sub-unit.

hutel_19g22: The new protein can find application in modulation of tissue-calcification, especially the uterus.

hutel_19h17: The new protein can find application in modulating the response of cells to oxysterols.

hutel_20b19: The novel protein seems to be a novel enzyme with sarcosine oxidase activity. The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

hutel_20g21: The novel protein seems to be a new ras inhibitor protein. The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

hutel_20h13: The novel protein is a new human alpha-adaptin. The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

hutel_20m11: The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

hutel_20m24: This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2. The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

hutel_22e12: The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

hutel_23e13: The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnologic production processes.

hutel_24j6: The new protein can find application in modulation of cell-cell-adhesion.

hutel_24h3: The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

Fetal Brain:

hfbr2_16c16: The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

hfbr2_23b21: The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

hfbr2_23b10: The new protein can find application in modulation of splicing.

hfbr2_2b5: The novel protein contains the typical (xxG)_n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain. The new protein can find application in modulation of connective tissue, bone and cartilage development and maintainance.

hfbr2_2c17: The new protein can find application in modulating/blocking G-protein-dependent pathways.

hfbr2_2d15: The new protein can find application in modulating early spermatogenesis.

hfbr2_2i17: The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

hfbr2_2k14: Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

hfbr_3c18: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

hfbr_3g8: The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

hfbr2_62b11: The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

hfbr2_62o17: The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins.

hfbr_6b24: The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

hfbr_72b18: The new protein can find application in modulating DNA repair and mutagenesis.

hfbr_78c4: The new protein can find application in modulating/blocking the response of cells to interferons.

hfbr_78k24: These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. The new protein can find application in modulation of protein stability/degradation in cells.

hfbr_82e4: The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

VARIANTS OF THE INVENTIVE DNA MOLECULES

Variants in General

"Variants," according to the invention, include DNA and/or protein molecules that resemble, structurally and/or functionally, those set forth in herein. Variants may be isolated from natural sources ("homologs"), may be entirely synthetic or may be based in part on both natural and synthetic approaches.

The section set forth below presents various structural and functional characteristics of molecules within the invention. Preferred molecules are characterized by a combination of one or more of these characteristics. For instance, some preferred molecules are described with reference to at least two structural characteristics, while others may be described with reference to at least one structural and at least one functional characteristic.

It will be recognized by the skilled artisan that structure ultimately defines function, *i.e.* the functions of the molecules described herein derives from the structures of those

molecules. Accordingly, the structural variants described below that bear the closest structural relationship (as variously defined below) to the inventive molecules are the variants that most likely will preserve biological function. This relationship between structure and function will guide the skilled artisan in identifying the preferred embodiments of the invention.

Splicing Variants

It is well-known that eukaryotic structural genes are comprised of both protein coding and non-coding portions. When the messenger RNA is transcribed from the DNA template, it contains introns, which are non-coding, and exons, which are coding. In order to form a translation competent mRNA, the introns must be "spliced" out of this initial pre mRNA.

Specific sequences within the pre mRNA represent "splice junctions" that direct the cellular splicing machinery to the appropriate position. The splice junctions are loosely conserved sequence regions of the pre mRNA, which almost invariably begin with GT and end with AG (DNA perspective). The 5' end of the splice junction typically contains about nine somewhat conserved residues, for example, C/AAGTA/GAGT. The 3' end usually contains a pyrimidine rich stretch of at least about 11 nucleotides, followed by NC/TAGG. Splicing occurs before the GT and after the AG. Mount, *Nucleic Acids Res.* 10:459-72 (1982).

Interestingly, exons often correspond to discrete functional domains of the protein product. The intron/exon arrangement thus creates a linear array of nucleotides which can be correlated to discrete, and often interchangeable, functional protein fragments. Go, *Nature* 291:90-92 (1981); Branden *et al.*, *EMBO J.* 3:1307-10 (1984). This linear arrangement creates the possibility of generating multiple different full length proteins by rearranging the order of the different functional portions in the array. For example, if a set of exons are arranged 1-2-3-4, where (-) represents the introns separating the exons, a splicing event need not simply produce 1234, but may produce 123, 134, 124 and so on. Production of different mRNA products in this way is commonly called "alternative splicing." Andreadis *et al.*, *Ann. Rev. Cell Biol.* 3:207-42 (1987).

Some of the present DNA molecules can be represented in modular fashion in terms of their coding regions. Essentially, these modules are exons (though each "exon" may in fact be made up of several exons), which may be combined in different ways to form a variety of

different DNA molecules, each encoding a different functional protein. Splicing variants are indicated below.

Degenerate Variants

One aspect of the present invention provides "degenerate variants" of the nucleic acid fragments of the present invention. A "degenerate variant" is a nucleotide fragment which differs from those of inventive molecules by nucleotide sequence, but due to the degeneracy of the genetic code, encodes an identical polypeptide sequence.

Given the known relationship between DNA sequences and the proteins they encode, degenerate variants typically are described by reference to this relationship. It is well known that the degeneracy of the genetic code results in many possible DNA sequences which encode a particular protein. Indeed, of the three bases which comprise an amino acid-encoding triplet, the third position, and often the second, almost always may vary. This fact alone allows for a class of variant DNA molecules which encode protein sequences identical to those disclosed herein, yet have about 30% sequence variation. In other words, the variant DNA molecules are about 70% identical to the inventive DNAs, having no additional or deleted sequences. Thus, one aspect of the invention provides degenerate variant DNA molecules encoding the inventive protein sequences.

In one embodiment, these variants have at least about 70% sequence identity with the DNA molecules described herein. In a preferred embodiment, these variants have at least about 80% sequence identity to the inventive molecules. In a more preferred embodiment these variants have at least about 90% sequence identity with the inventive molecules.

Conservative Amino Acid Variants

Variants according to the invention also may be made that conserve the overall molecular structure of the encoded proteins. Given the properties of the individual amino acids comprising the disclosed protein products, some rational substitutions will be recognized by the skilled worker. Amino acid substitutions, *i.e.* "conservative substitutions," may be made, for instance, on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example: (a) nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; (b) polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine;

(c) positively charged (basic) amino acids include arginine, lysine, and histidine; and (d) negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Substitutions typically may be made within groups (a)-(d). In addition, glycine and proline may be substituted for one another based on their ability to disrupt α -helices. Similarly, certain amino acids, such as alanine, cysteine, leucine, methionine, glutamic acid, glutamine, histidine and lysine are more commonly found in α -helices, while valine, isoleucine, phenylalanine, tyrosine, tryptophan and threonine are more commonly found in β -pleated sheets. Glycine, serine, aspartic acid, asparagine, and proline are commonly found in turns. Some preferred substitutions may be made among the following groups: (i) S and T; (ii) P and G; and (iii) A, V, L and I. Given the known genetic code, and recombinant and synthetic DNA techniques, the skilled scientist readily can construct DNAs encoding the conservative amino acid variants.

As used herein, "sequence identity" between two polypeptide sequences indicates the percentage of amino acids that are identical between the sequences. "Sequence similarity" indicates the percentage of amino acids that either are identical or that represent conservative amino acid substitutions.

Functionally Equivalent Variants

Yet another class of DNA variants within the scope of the invention may be described with reference to the product they encode. As shown below, some of the inventive DNA molecules encode a protein having a degree of homology with known proteins, or protein domains. It is expected, therefore, that they will have some or all of the requisite functional features of such molecules. These "functionally equivalent variants" products are characterized by the fact that they are functionally equivalent, with respect to biological activity, to certain known molecules.

The instant invention provides information on common structural motifs, including consensus sequences that will guide the artisan in constructing functionally equivalent variants. It will be understood that the motifs, identified for each inventive protein, may be modified within the identified consensus sequences. Thus, the invention contemplates the proteins disclosed herein that contain variability in the consensus sequences identified, and the invention further contemplates the full range of nucleic acids encoding them, and the complements of those nucleic acids.

Hybridizing Variants

DNA variants within the invention also may be described by reference to their physical properties in hybridization. One skilled in the field will recognize that DNA can be used to identify its complement and, since DNA is double stranded, its equivalent or homolog, using nucleic acid hybridization techniques. It will also be recognized that hybridization can occur with less than 100% complementarity. However, given appropriate choice of conditions, hybridization techniques can be used to differentiate among DNA sequences based on their structural relatedness to a particular probe. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, MOLECULAR CLONING, A LABORATORY MANUAL, Cold Spring Harbor Press, N.Y.; and Ausubel *et al.*, 1989, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Green Publishing Associates and Wiley Interscience, N.Y.

Structural relatedness between two polynucleotide sequences can be expressed as a function of "stringency" of the conditions under which the two sequences will hybridize with one another. As used herein, the term "stringency" refers to the extent that the conditions disfavor hybridization. Stringent conditions strongly disfavor hybridization, and only the most structurally related molecules will hybridize to one another under such conditions. Conversely, non-stringent conditions favor hybridization of molecules displaying a lesser degree of structural relatedness. Hybridization stringency, therefore, directly correlates with the structural relationships of two nucleic acid sequences. The following relationships are useful in correlating hybridization and relatedness (where T_m is the melting temperature of a nucleic acid duplex):

- a. $T_m = 69.3 + 0.41(G+C)\%$
- b. The T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatched base pairs.
- c. $(T_m)_{\mu 2} - (T_m)_{\mu 1} = 18.5 \log_{10} \mu 2 / \mu 1$
where $\mu 1$ and $\mu 2$ are the ionic strengths of two solutions.

Hybridization stringency is a function of many factors, including overall DNA concentration, ionic strength, temperature, probe size and the presence of agents which disrupt hydrogen bonding. Factors promoting hybridization include high DNA

concentrations, high ionic strengths, low temperatures, longer probe size and the absence of agents that disrupt hydrogen bonding.

Hybridization usually is done in two stages. First, in the "binding" stage, the probe is bound to the target under conditions favoring hybridization. Stringency is usually controlled at this stage by altering the temperature. For high stringency, the temperature is usually between 65°C and 70°C, unless short (<20 nt) oligonucleotide probes are used. A representative hybridization solution comprises 6X SSC, 0.5% SDS, 5X Denhardt's solution and 100µg of non-specific carrier DNA. See Ausubel *et al.*, *supra*, section 2.9, supplement 27 (1994). Of course many different, yet functionally equivalent, buffer conditions are known. Where the degree of relatedness is lower, a lower temperature may be chosen. Low stringency binding temperatures are between about 25°C and 40°C. Medium stringency is between at least about 40°C to less than about 65°C. High stringency is at least about 65°C.

Second, the excess probe is removed by washing. It is at this stage that more stringent conditions usually are applied. Hence, it is this "washing" stage that is most important in determining relatedness via hybridization. Washing solutions typically contain lower salt concentrations. One exemplary medium stringency solution contains 2X SSC and 0.1% SDS. A high stringency wash solution contains the equivalent (in ionic strength) of less than about 0.2X SSC, with a preferred stringent solution containing about 0.1X SSC. The temperatures associated with various stringencies are the same as discussed above for "binding." The washing solution also typically is replaced a number of times during washing. For example, typical high stringency washing conditions comprise washing twice for 30 minutes at 55° C. and three times for 15 minutes at 60° C.

The present invention includes nucleic acid molecules that hybridize to the inventive molecules under high stringency binding and washing conditions. More preferred molecules (from an mRNA perspective) are those that are at least 50 % of the length of any one of those depicted in below. Particularly preferred molecules are at least 75 % of the length of those molecules.

Substitutions, Insertions, Additions and Deletions

In a general sense, the preferred DNA variants of the invention are those that retain the closest relationship, as described by "sequence identity" to the inventive DNA molecules. According to another aspect of the invention, therefore, substitutions, insertions, additions and deletions of defined properties are contemplated. It will be recognized that sequence

identity between two polynucleotide sequences, as defined herein, generally is determined with reference to the protein coding region of the sequences. Thus, this definition does not at all limit the amount of DNA, such as vector DNA, that may be attached to the molecules described herein. Preferred DNA sequence variants include molecules encoding proteins sharing some or all of any relevant biological activity of the native molecule.

In creating these variants, the skilled worker will be guided by reference to the protein structure. First, insertions and deletions in any recognized functional domain, above, generally should be avoided, except as noted below in the section entitled "Proteins," where this domain is discussed in detail. Alterations in such domains usually will be limited to conservative amino acid substitutions. In addition, where insertions and deletions are desired, this may be accomplished at the N- and/or C-terminus of the protein molecule (or the corresponding coding regions of the DNA). If insertions or deletions are made within the protein, deletions of major structural features usually should be avoided. Thus, a preferred place to make insertion or deletion variants is in non-structural regions, such as linker regions between two alpha helices.

"Substitutions" generally refer to alterations in the DNA sequence which do not change its overall length, but only alter one or more nucleotide positions, substituting one for another in the common sense of the word. One class of preferred substitutions, "degenerate substitutions," are those that do not alter the encoded amino acid sequence. Some substitutions retains 50%, 55%, 60% or 65% identity. Preferred substitutions retain at least about 70% identity, more preferably at least 70% or 75% identity, with the inventive DNAs. Some more preferred molecules have at least about 80% identity, more preferably at least 80% or 85% identity. Particularly preferred DNAs share at least about 90% identity, more preferably at least 90% or 95% identity.

"Insertions," unlike substitutions, alter the overall length of the DNA molecule, and thus sometimes the encoded protein. Insertions add extra nucleotides to the interior (not the 5' or 3' ends) of the subject DNAs. Preferred insertions are made with reference to the protein sequence encoded by the DNA. Thus, it is most preferred to provide an insertion in the DNA at a location that corresponds to an area of the encoded protein which lacks structure. For instance, it typically would not be beneficial, if the preservation of biological activity is desired, to provide an insertion within an alpha-helical region or a beta-pleated sheet. Accordingly, non-structural areas, such as those containing helix-breaking glycines

and proline residues, are most preferred sites of insertion. Other preferred sites of insertion are the splice sites, which are indicated above in the description of the inventive DNA molecules.

While the optimal size of insertions will vary depending upon the site of insertion and its effect on the overall conformation of the encoded protein, some general guides are useful. Generally, the total insertions (irrespective of their number) should not add more than about 30% (or preferably not more than 30%) to the overall size of the encoded protein. More preferably, the insertion adds less than about 10-20% (yet more preferably 10-20%) in size, with less than about 10% being most preferred. The number of insertions is limited only by the number of suitable insertions sites, and secondarily by the foregoing size preferences.

"Additions," like insertions, also add to the overall size of the DNA molecule, and usually the encoded protein. However, instead of being made within the molecule, they are made on the 5' or 3' end, usually corresponding to the N- or C- terminus of the encoded protein. Unlike deletions, additions are not very size-dependent. Indeed, additions may be of virtually any size. Preferred additions, however, do not exceed about 100% of the size of the native molecule. More preferably, they add less than about 60 to 30% to the overall size, with less than about 30% being most preferred.

"Deletions" diminish the overall size of the DNA and, therefore, also reduce the size of the protein encoded by that DNA. Deletions may be made from either end of the molecule or internal to it. Typical preferred deletions remove discrete structural features of the encoded protein. For example, some deletions will comprise the deletion of one or more exons which may define a structural feature. Preferred deletions remove less than about 30% of the size of the subject molecule. More preferred deletions remove less than about 20% and most preferred deletions remove less than about 10%.

Computer-Defined Variants and Definition of "Sequence Identity"

In general, both the DNA and protein molecules of the invention can be defined with reference to "sequence identity." As used herein, "sequence identity" refers to a comparison made between two molecules using, for example, the standard Smith-Waterman algorithm that is well known in the art.

Some molecules have at least about 50%, 55% or 60% identity. Preferred molecules are those having at least about 65% sequence identity, more preferably at least 65% or 70% sequence identity. Other preferred molecules have at least about 80%, more preferably at

least 80% or 85%, sequence identity. Particularly preferred molecules have at least about 90% sequence identity, more preferably at least 90% sequence identity. Most preferred molecules have at least about 95%, more preferably at least 95%, sequence identity. As used herein, two nucleic acid molecules or proteins are said to "share significant sequence identity" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) identity.

"Sequence identity" is defined herein with reference the Blast 2 algorithm, which is available at the NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>), using default parameters.

References pertaining to this algorithm include: those found at

http://www.ncbi.nlm.nih.gov/BLAST/blast_references.html; Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410; Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3:266-272; Madden, T.L., Tatusov, R.L. & Zhang, J. (1996) "Applications of network BLAST server" Meth. Enzymol. 266:131-141; Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402; and Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649-656.

METHODS OF MAKING VARIANTS

It will be recognized that variants of the inventive molecules can be constructed in several different ways. For example, they may be constructed as completely synthetic DNAs. Methods of efficiently synthesizing oligonucleotides in the range of 20 to about 150 nucleotides are widely available. See Ausubel *et al.*, *supra*, section 2.11, Supplement 21 (1993). Overlapping oligonucleotides may be synthesized and assembled in a fashion first reported by Khorana *et al.*, J. Mol. Biol. 72:209-217 (1971); see also Ausubel *et al.*, Section 8.2. The synthetic DNAs are designed with convenient restriction sites engineered at the 5' and 3' ends of the gene to facilitate cloning into an appropriate vector.

An alternative method of generating variants is to start with one of the inventive DNAs and then to conduct site-directed mutagenesis. See Ausubel *et al.*, *supra*, chapter 8, Supplement 37 (1997). In a typical method, a target DNA is cloned into a single-stranded

DNA bacteriophage vehicle. Single-stranded DNA is isolated and hybridized with a oligonucleotide containing the desired nucleotide alteration(s). The complementary strand is synthesized and the double stranded phage is introduced into a host. Some of the resulting progeny will contain the desired mutant, which can be confirmed using DNA sequencing. In addition, various methods are available that increase the probability that the progeny phage will be the desired mutant. These methods are well known to those in the field and kits are commercially available for generating such mutants.

ISOLATING HOMOLOGS

Methods

By using the sequences disclosed herein as probes or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs. "Homologs" are essentially naturally-occurring variants and include allelic, species-specific and tissue-specific variants.

Region-specific primers or probes derived from the nucleotide sequence(s) provided can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog using known methods (Innis *et al.*, *PCR Protocols*, Academic Press, San Diego, CA (1990)). Such an application is useful in diagnostic methods, as described in more detail below, as well as in preparing full-length DNAs from various sources. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. As a general guide, the formula $3(G+C) + 2(A+T) = ^\circ\text{C}$, is useful.

When using primers derived from the inventive sequences, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C), only sequences with greater than 75% sequence identity to the primer will be amplified. By employing lower stringency conditions (*e.g.*, annealing at 35-37°C), sequences which have greater than 40-50% sequence identity to the primer also will be amplified.

The PCR product may be subcloned and sequenced to confirm that it indeed displays the expected sequence identity. The PCR fragment may then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled

and used to screen a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

PCR technology may also be utilized to isolate full length cDNA sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of cloning strategies which may be used, see e.g., Sambrook et al., 1989, *supra*.

When using DNA probes derived from the inventive sequences for colony/plaque hybridization, one skilled in the art will recognize that by employing medium to high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC), sequences having regions with greater than 90% sequence identity to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in SSPC), sequences having regions with greater than 35-45% sequence identity to the probe will be obtained.

Suitably, genomic or cDNA libraries can be constructed and screened in accord with the previous paragraph. The libraries should be derived from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. The clone containing the homolog may then be purified through methods routinely practiced in the art, and subjected to sequence analysis.

Additionally, an expression library can be constructed utilizing DNA isolated from or cDNA synthesized from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. In this manner, clones may be induced and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal gene product, as described herein. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, *ANTIBODIES: A LABORATORY MANUAL*, Cold Spring Harbor Press, Cold Spring Harbor Press.)

Human Homologs

Any organism or tissue can be used as the source for homologs of the present invention so long as the organism or tissue naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs is human.

PROTEINS OF THE INVENTION

One class of proteins included within the invention is encoded by the inventive DNA molecules presented. Other proteins according to the invention are those encoded by the DNA variants described above. As noted, these variants are designed with the encoded proteins in mind.

A preferred class of protein fragments includes those fragments which retain any biological activity. These molecules share functional features common the family of proteins, although these characteristics may vary in degree.

According to one aspect of the invention fragments of the inventive proteins are contemplated. Some preferred fragments are those which are capable of eliciting an immune response. Generally these "antigenic" fragments will be from about five amino acids in length to about fifty amino acids in length. Some preferred antigenic fragments are from five to about twenty amino acids long. "Antigenic" response may refer to a T cell response, a B cell response or a response by cells of the macrophage/monocyte lineages. In most cases, however, it will refer to the immune response involved in the generation of antibodies. In other words, the relevant immune response is that of helper T cells and/or B cells. These preferred molecules comprise one or more T cell and /or B cell epitopes.

ANTIBODIES OF THE INVENTION

Antibodies raised against the proteins and protein fragments of the invention also are contemplated by the invention. Described below are antibody products and methods for producing antibodies capable of specifically recognizing one or more epitopes of the presently described proteins and their derivatives.

Antibodies include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies including single chain Fv (scFv) fragments, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, epitope-binding fragments, and humanized forms of any of the above.

As known to one in the art, these antibodies may be used, for example, in the detection of a target protein in a biological sample. They also may be utilized as part of treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels or for the presence of abnormal forms of the such proteins.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., *J. Immunol. Methods* 35:1-21 (1980); Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today* 4:72 (1983); Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985), pp. 77-96). Antibodies may also be generated by the known techniques of phage display and *in vitro* immunization.

Polyclonal Antibodies

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as an inventive protein or an antigenic derivative thereof.

Polyclonal antiserum, containing antibodies to heterogeneous epitopes of a single protein, can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified, as known in the art, to enhance immunogenicity. Immunization methods include subcutaneous or intraperitoneal injection of the polypeptide.

Effective polyclonal antibody production is affected by many factors related both to the antigen and to the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and/or adjuvant. In addition, host animal response may vary with site of inoculation. Both inadequate or excessive doses of antigen may result in low titer antisera. In general, however, small doses (high ng to low µg levels) of antigen administered at multiple intradermal sites appears to be most reliable. Host animals may include but are not limited to rabbits, mice, chickens and rats, to name but a few. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al., *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

The protein immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin β -galactosidase) or through the inclusion of an adjuvant during immunization. Adjuvants include Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*.

Booster injections can be given at regular intervals, with at least one usually being required for optimal antibody production. The antiserum may be harvested when the antibody titer begins to fall. Titer may be determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen. See, for example, Ouchterlony *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μ M). The antiserum may be purified by affinity chromatography using the immobilized immunogen carried on a solid support. Such methods of affinity chromatography are well known in the art.

Affinity of the antisera for the antigen may be determined by preparing competitive binding curves, as described, for example, by Fisher, Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D.C. (1980).

In addition to using protein as the immunogen, DNA molecules may be used directly. In this manner, a DNA encoding the protein immunogen is administered. Boosting and harvesting is done in a manner analogous to that detailed above. Yet another method of producing antibodies entails immunizing chickens and harvesting the antibodies from their eggs.

Monoclonal Antibodies

Monoclonal antibodies (MAbs), are homogeneous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture or *in vivo*. MAbs may be produced

by making hybridomas which are immortalized cells capable of secreting a specific monoclonal antibody.

Monoclonal antibodies to any of the proteins, peptides and epitopes thereof described herein can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495-497 (1975) (and U.S. Patent No. 4,376,110) or modifications of the methods thereof, such as the human B-cell hybridoma technique (Kosbor *et al.*, 1983, *Immunology Today* 4:72; Cole *et al.*, 1983, *Proc. Natl. Acad. Sci. USA* 80: 2026-2030), and the EBV-hybridoma technique (Cole *et al.*, 1985, *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96).

In one method a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen are isolated.

The spleen cells are fused, typically using polyethylene glycol, with mouse myeloma cells, such as SP2/0-Ag14 myeloma cells. The excess, unfused cells are destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted, and aliquots are plated to microliter plates where growth is continued.

Antibody-producing clones (hybridomas) are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures. These include ELISA, as originally described by Engvall, *Meth. Enzymol.* 70:419 (1980), western blot analysis, radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124 (1988)) and modified methods thereof.

Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* *BASIC METHODS IN MOLECULAR BIOLOGY*, Elsevier, New York. Section 21-2 (1989). The hybridoma clones may be cultivated *in vitro* or *in vivo*, for instance as ascites. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production. Alternatively, hybridoma culture in hollow fiber bioreactors provides a continuous high yield source of monoclonal antibodies.

The antibody class and subclass may be determined using procedures known in the art (Campbell, A.M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

MAbs may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. Methods of purifying monoclonal antibodies are well known in the art.

Antibody Derivatives and Fragments

Fragments or derivatives of antibodies include any portion of the antibody which is capable of binding the target antigen, or a specific portion thereof. Antibody derivatives include poly-specific (*e.g.*, bi-specific) antibodies, which contain binding sites specific for two or more different epitopes. These epitopes may be from the same or different inventive molecules or one or more epitope may be from a molecule not specifically disclosed here.

Antibody fragments specifically include $F(ab')_2$, Fab, Fab' and Fv fragments. These can be generated from any class of antibody, but typically are made from IgG or IgM. They may be made by conventional recombinant DNA techniques or, using the classical method, by proteolytic digestion with papain or pepsin. See CURRENT PROTOCOLS IN IMMUNOLOGY, chapter 2, Coligan *et al.*, eds., (John Wiley & Sons 1991-92).

$F(ab')_2$ fragments are typically about 110 kDa (IgG) or about 150 kDa (IgM) and contain two antigen-binding regions, joined at the hinge by disulfide bond(s). Virtually all, if not all, of the Fc is absent in these fragments. Fab' fragments are typically about 55 kDa (IgG) or about 75 kDa (IgM) and can be formed, for example, by reducing the disulfide bond(s) of an $F(ab')_2$ fragment. The resulting free sulfhydryl group(s) may be used to conveniently conjugate Fab' fragments to other molecules, such as detection reagents (*e.g.*, enzymes).

Fab fragments are monovalent and usually are about 50 kDa (from any source). Fab fragments include the light (L) and heavy (H) chain, variable (V_L and V_H , respectively) and constant (C_L C_H , respectively) regions of the antigen-binding portion of the antibody. The H and L portions are linked by an intramolecular disulfide bridge.

Fv fragments are typically about 25 kDa (regardless of source) and contain the variable regions of both the light and heavy chains (V_L and V_H , respectively). Usually, the V_L and V_H chains are held together only by non-covalent interactions and, thus, they readily dissociate. They do, however, have the advantage of small size and they retain the same binding properties of the larger Fab fragments. Accordingly, methods have been developed to crosslink the V_L and V_H chains, using, for example, glutaraldehyde (or other chemical crosslinkers), intermolecular disulfide bonds (by incorporation of cysteines) and peptide linkers. The resulting Fv is now a single chain (*i.e.*, SCFv).

Other antibody derivatives include single chain antibodies (U.S. Patent 4,946,778; Bird, *Science* 242:423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85:5879-5883 (1988); and Ward *et al.*, *Nature* 334:544-546 (1989)). Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain FV (SCFv).

One preferred method involves the generation of scFvs by recombinant methods, which allows the generation of Fvs with new specificities by mixing and matching variable chains from different antibody sources. In a typical method, a recombinant vector would be provided which comprises the appropriate regulatory elements driving expression of a cassette region. The cassette region would contain a DNA encoding a peptide linker, with convenient sites at both the 5' and 3' ends of the linker for generating fusion proteins. The DNA encoding a variable region(s) of interest may be cloned in the vector to form fusion proteins with the linker, thus generating an scFv.

In an exemplary alternative approach, DNAs encoding two Fvs may be ligated to the DNA encoding the linker, and the resulting tripartite fusion may be ligated directly into a conventional expression vector. The scFv DNAs generated any of these methods may be expressed in prokaryotic or eukaryotic cells, depending on the vector chosen.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')₂ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab)₂ fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, 1989, *Science*, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Derivatives also include "chimeric antibodies" (Morrison *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851-6855 (1984); Neuberger *et al.*, *Nature*, 312:604-608 (1984); Takeda *et al.*, *Nature*, 314:452-454 (1985)). These chimeras are made by splicing the DNA encoding a mouse antibody molecule of appropriate specificity with, for instance, DNA encoding a human antibody molecule of appropriate specificity. Thus, a chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. These are also known sometimes as "humanized" antibodies and they offer the added

advantage of at least partial shielding from the human immune system. They are, therefore, particularly useful in therapeutic *in vivo* applications.

Labeled Antibodies

The present invention further provides the above-described antibodies in detectably labeled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example see (Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer *et al.*, *Meth. Enzym.* 62:308 (1979); Engval *et al.*, *Immunol.* 109:129 (1972); Goding, *J. Immunol. Meth.* 13:215 (1976)). The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* diagnostic assays.

Immobilized Antibodies

The foregoing antibodies also may be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby *et al.*, *Meth. Enzym.* 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immunoaffinity purification of the proteins of the present invention.

THERAPEUTIC AND DIAGNOSTIC COMPOSITIONS

The proteins, antibodies and polynucleotides of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e.g., human serum albumin, are described, for example, in *Remington's Pharmaceutical Sciences* (16th ed., Osol, A., Ed., Mack, Easton PA (1980)). In order to form a pharmaceutically acceptable composition suitable for effective administration,

such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvate may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound. For buccal administration the composition may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.* gelatin for

use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

RECOMBINANT CONSTRUCTS AND EXPRESSION

The present invention further provides recombinant DNA constructs comprising one or more of the nucleotide sequences of the present invention. The recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a DNA or DNA fragment, typically bearing an open reading frame, is inserted, in either orientation.

The gene products encoded by the subject DNAs may be produced by recombinant DNA technology using techniques well known in the art. See, for example, the techniques described in Sambrook et al., 1989, *supra*, and Ausubel et al., 1989, *supra*. Alternatively, the DNA sequences may be chemically synthesized using, for example, synthesizers. See, for

example, the techniques described in OLIGONUCLEOTIDE SYNTHESIS, 1984, Gait, ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety. They may be assembled from fragments and short oligonucleotide linkers, or from a series of oligonucleotides. They are preferably made by RT-PCR methods. The resulting synthetic gene is capable of being expressed in a recombinant vector.

In some cases the recombinant constructs will be expression vectors, which are capable of expressing the RNA and/or protein products of the encoded DNA(s). Thus, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the open reading frame (ORF). The vector may further comprise a selectable marker sequence.

Specific initiation signals may also be required for efficient translation of inserted target gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where a target DNA includes its own initiation codon and adjacent sequences is inserted into the appropriate expression vector, no additional translation control signals may be needed. However, in cases where only a portion of an ORF is used, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire target. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner *et al.*, *Methods in Enzymol.* 153:516-544 (1987)). Some appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U.S. Patent No. 5,082,767.

The present invention further provides host cells containing at least one of the DNAs of the present invention. The host cell can be virtually any cell for which expression vectors are available. It may be, for example, a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic

cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis *et al.*, *Basic Methods in Molecular Biology* (1986)).

A wide variety of expression systems are available, such as: yeast (*e.g.* *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing the target DNA; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the target DNA sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.* Ti plasmid) containing target DNA coding sequences; or mammalian cell systems (*e.g.* COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

Depending on the system chosen, the resulting product may differ. For example, proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

Vectors

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting selection of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequence, and in one aspect of the invention, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal or C-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Bacterial Expression

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, if desirable, to provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may, also be employed as a matter of choice.

Bacterial vectors may be, for example, bacteriophage-, plasmid- or cosmid-based. These vectors can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids typically containing elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, GEM 1 (Promega Biotec, Madison, WI, USA), pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pKK232-8, pDR540, and pRIT5 (Pharmacia).

These "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Bacterial promoters include lac, T3, T7, lambda P_R or P_L, trp, and ara.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is derepressed/induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, *EMBO J.* 2:1791), in which the coding sequence may be ligated into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye et al. 1985, *Nucleic Acids*

Res. 13:3101-3109; Van Heeke *et al.*, 1989, *J. Biol. Chem.* 264:5503-5509); pET vectors, Studier *et al.*, *Methods in Enzymology* 185: 60-89 (Academic Press 1990); and the like.

Moreover, pGEX vectors may be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and easily can be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In a one embodiment, full length cDNA sequences are appended with in-frame *Bam*HI sites at the amino terminus and *Eco*RI sites at the carboxyl terminus using standard PCR methodologies (Innis *et al.*, 1990, *supra*) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive labeling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.* 1985, *EMBO J.* 4: 1075; Zabeau and Stanley, 1982, *EMBO J.* 1:1217).

Eukaryotic Expression

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Mammalian promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Exemplary mammalian vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). Selectable markers include CAT (chloramphenicol transferase).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the coding sequence of interest

may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a target protein in infected hosts. (*E.g.*, See Logan *et al.*, 1984, *Proc. Natl. Acad. Sci. USA* 81:3655-3659).

In one embodiment, cDNA sequences encoding the full-length open reading frames are ligated into pCMV β replacing the β -galactosidase gene such that cDNA expression is driven by the CMV promoter (Alam, 1990, *Anal. Biochem.* 188: 245-254; MacGregor *et al.*, 1989, *Nucl. Acids Res.* 17: 2365; Norton *et al.* 1985, *Mol. Cell. Biol.* 5: 281).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins.

Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins in eukaryotic cells, stable expression is preferred. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, *etc.*), and a selectable marker.

Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the target protein. Such engineered cell lines may be

particularly useful in screening and evaluation of compounds that affect the endogenous activity of the protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, *et al.*, *Cell* 11:223 (1977)), hypoxanthine-guanine phosphoribosyltransferase (Szybalska *et al.*, *Proc. Natl. Acad. Sci. USA* 48:2026 (1962)), and adenine phosphoribosyltransferase (Lowy, *et al.*, *Cell* 22:817 (1980)) genes can be employed in tk⁻, hgp^r or ap^r cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, *et al.*, *Proc. Natl. Acad. Sci. USA* 77:3567 (1980)); O'Hare, *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2072 (1981)); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, *et al.*, 1981, *J. Mol. Biol.* 150:1); and hydro, which confers resistance to hygromycin (Santerre, *et al.*, 1984, *Gene* 30:147) genes.

An alternative fusion protein system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 8972-8976 (1991)). In this system, the gene of interest is subcloned into a vaccinia-based plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni²⁺ nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The target coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of a target gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (*E.g.*, see Smith *et al.*, 1983, *J. Virol.* 46: 584; Smith, U.S. Patent No. 4,215,051).

While the present proteins can be expressed in recombinant systems, as described above, cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

Purification of Recombinant Proteins

Recombinant proteins produced may be isolated by host cell lysis. This may be followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, like lysozyme and chelators.

If inclusion bodies are formed in bacterial systems, they may be extracted from cell pellets using, for example, detergents, reducing agents, salts, urea, guanidinium chloride and extremes of pH (*e.g.* <4 or >10). If denaturation occurs, protein refolding steps (*e.g.*, dialysis) can be used, as necessary, in completing configuration of the mature protein. If disulfide bridges are present in the native protein, they may be reoxidized using known methods.

By way of specific non-limiting example, the recombinant bacterial cells, for example *E. coli*, are grown in any of a number of suitable media, for example LB, and the expression of the recombinant protein induced by adding IPTG (*e.g.*, *lac* operator-promoter) to the media or switching incubation to a higher temperature (*e.g.*, λ cl^{857}). After culturing the bacteria for a further period of between 2 and 24 hours, the cells are collected by centrifugation and washed to remove residual media. The bacterial cells are then lysed, for example, by disruption in a cell homogenizer and centrifuged to separate the cell membranes from the soluble cell components. If the protein aggregates into inclusion bodies, this centrifugation can be performed under conditions whereby the dense inclusion bodies are selectively enriched by incorporation of sugars such as sucrose into the buffer and centrifugation at a selective speed. The inclusion bodies can then be washed in any of several solutions to remove some of the contaminating host proteins, then solubilized in solutions containing high concentrations of urea (*e.g.* 8M) or chaotropic agents such as guanidinium hydrochloride in the presence of reducing agents such as β -mercaptoethanol or DTT (dithiothreitol).

At this stage it may be advantageous to incubate the protein for several hours under conditions suitable for the protein to undergo a refolding process into a conformation which

more closely resembles that of the native protein. Such conditions generally include low protein concentrations less than 500 µg/ml), low levels of reducing agent, concentrations of urea less than 2 M and often the presence of reagents such as a mixture of reduced and oxidized glutathione which facilitate the interchange of disulphide bonds within the protein molecule. The refolding process can be monitored, for example, by SDS-PAGE or with antibodies which are specific for the native molecule. Following refolding, the protein can then be purified further and separated from the refolding mixture by chromatography on any of several supports including ion exchange resins, gel permeation resins or on a variety of affinity columns.

Labeling Proteins

When used as a component in assay systems such as those described, below, the target protein may be labeled, either directly or indirectly, to facilitate detection of the present *res*-like molecules either *in vitro* or *in vivo*. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ¹²⁵I; enzyme labeling systems that generate a detectable colorimetric signal or light when exposed to substrate; and fluorescent labels.

Where recombinant DNA technology is used for protein production the, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection. These fusion proteins may, for example, add amino acids which facilitate further chemical modification. They also may add a functional moiety, such as an enzyme, which directly facilitates detection.

TRANSGENIC ANIMALS

The invention further contemplates animal models for studying the function of the present molecules and for overproducing the protein products. The disclosed DNA sequences may be used in conjunction with techniques for producing transgenic animals that are well known to those of skill in the art.

To prepare transgenic animals, target gene sequences may for example be introduced into, and overexpressed in, the genome of the animal of interest, or, if endogenous target gene sequences are present, they may either be overexpressed or, alternatively, be disrupted in order to underexpress or inactivate target gene expression, such as described for the disruption of apoE in mice (Plum *et al.*, *Cell* 71: 343-353 (1992)).

In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence which is capable of driving gene expression in the animal and cell type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation.

For underexpression of an endogenous target gene sequence, such a sequence may be isolated and engineered such that when reintroduced into the genome of the animal of interest, the endogenous target gene alleles will be inactivated. Preferably, the engineered target gene sequence is introduced via gene targeting such that the endogenous target sequence is disrupted upon integration of the engineered target gene sequence into the animal's genome.

Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate cardiovascular disease animal models. Goats, cows and sheep are particularly preferred for producing protein *in vivo*.

Any technique known in the art may be used to introduce a target gene transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe *et al.*, U.S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82:6148-6152 (1985)); gene targeting in embryonic stem cells (Thompson *et al.*, *Cell* 56:313-321 (1989)); electroporation of embryos (Lo, *Mol. Cell. Biol.* 3:1803-1814 (1983)); and sperm-mediated gene transfer (Lavitrano *et al.*, *Cell* 57:717-723 (1989)); *etc.* For a review of such techniques, see Gordon, Transgenic Animals, *Intl. Rev. Cytol.* 115:171-229 (1989).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching

of Lasko et al. (Lasko *et al.*, *Proc. Natl. Acad. Sci. USA* 89:3232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the target gene be integrated into the chromosomal site of the endogenous target gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous target gene of interest are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous target gene.

The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene of interest in only that cell type, by following, for example, the teaching of Gu *et al.* *Science* 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant target gene and protein may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include but are not limited to Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and RT-PCR. Samples of target gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the target gene transgene gene product of interest.

The transgenic animals that express target gene mRNA or target gene transgene peptide (detected immunocytochemically, using antibodies directed against the target gene product's epitopes) at easily detectable levels should then be further evaluated to identify those animals which display characteristic increased susceptibility to carcinogenesis. Additionally, specific cell types within the transgenic animals may be analyzed and assayed *in vitro* for cellular phenotypes characteristic of mutant phenotype.

Once target gene transgenic founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include but are not limited to: outbreeding of founder animals with more

than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound target gene transgenics that express the target gene transgene of interest at higher levels because of the effects of additive expression of each target gene transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order both to augment expression and eliminate the possible need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; breeding animals to different inbred genetic backgrounds so as to examine effects of modifying alleles on expression of the target gene transgene and the possible development of carcinogenesis. One such approach is to cross the target gene transgenic founder animals with a wild type strain to produce an F1 generation that exhibits increased susceptibility to carcinogenesis. The F1 generation may then be inbred in order to develop a homozygous line, if it is found that homozygous target gene transgenic animals are viable.

Methods of generating "knockout" mice using homologous recombination in embryonic stem cells are well known in the art. Suitable methods are described, for example, in Mansour *et al.*, *Nature*, 336:348 (1988); Zijlstra *et al.*, *Nature*, 342:435 (1989) and 344:742 (1990); and Hasty *et al.*, *Nature*, 350:243 (1991). This genomic DNA can be obtained by conventional methods using the cDNA sequence as a probe in a commercially-available genomic DNA library.

Briefly, a genomic fragment is cleaved with a restriction endonuclease and a heterologous cassette containing a neomycin-resistance gene is inserted at the cleavage site. A suitable cassette is the GTI-II *neo* cassette described by Lufkin *et al.*, *Cell* 66:1105 (1991). The modified genomic fragment is cloned into a suitable targeting vector that is introduced into murine embryonic stem cells by electroporation. Cells that have undergone homologous recombination (and hence disruption of the gene) are selected by resistance to G418, and used to generate chimeric mice using well known methods. See Lufkin *et al.*, *supra*. Traditional breeding methods then can be used to generate mice that are homozygous for the disrupted gene.

The phenotype of mice that are homozygous for the mutation then can be studied to provide insights into the role of the protein in, for example, carcinogenesis. These mice also can be used as models for developing new treatments for cancers. If this mutation is lethal in

homozygous mice (for example during embryogenesis) heterozygous mice, which express only half the amount of the protein can also be studied.

GENE THERAPY APPLICATIONS

When mutations in the inventive protein, or in the elements controlling expression of that protein, are found to be associated with a malignant phenotype, control of cellular proliferation can be restored by gene therapy methods. For example, overexpression of the protein can be counteracted by concurrent expression of an antisense molecule that binds to and inhibits expression of the mRNA encoding the protein. Alternatively, overexpression can be inhibited in an analogous manner using a ribozyme that cleaves the mRNA. In another embodiment, where expression of a mutated protein induces the malignant phenotype, concomitant expression of the non-mutated molecule via introduction of an exogenous gene may be used. Methods of using antisense and ribozyme technology to control gene expression, or of gene therapy methods for expression of an exogenous gene in this manner are well known in the art.

Each of these methods requires a system for introducing a vector into the cells containing the mutated gene. The vector encodes either an antisense or ribozyme transcript of the inventive protein. The construction of a suitable vector can be achieved by any of the methods well-known in the art for the insertion of exogenous DNA into a vector. *See, e.g.,* Sambrook *et al.*, *Molecular Cloning* (Cold Spring Harbor Press 2d ed. 1989), which is incorporated herein by reference. In addition, the prior art teaches various methods of introducing exogenous genes into cells *in vivo*. *See* Rosenberg *et al.*, *Science* 242:1575-1578 (1988) and Wolff *et al.*, *PNAS* 86:9011-9014 (1989), which are incorporated herein by reference. The routes of delivery include systemic administration and administration *in situ*. Well-known techniques include systemic administration with cationic liposomes, and administration *in situ* with viral vectors. Any one of the gene delivery methodologies described in the prior art is suitable for the introduction of a recombinant vector containing an inventive gene according to the invention into a MTX-resistant, transport-deficient cancer cell. A listing of present-day vectors suitable for the purpose of this invention is set forth in Hodgson, *Bio/Technology* 13: 222 (1995), which is incorporated by reference.

For example, liposome-mediated gene transfer is a suitable method for the introduction of a recombinant vector containing an inventive gene according to the invention

into a MTX-resistant, transport-deficient cancer cell. The use of a cationic liposome, such as DC-Chol/DOPE liposome, has been widely documented as an appropriate vehicle to deliver DNA to a wide range of tissues through intravenous injection of DNA/cationic liposome complexes. See Caplen *et al.*, *Nature Med.* 1:39-46 (1995) and Zhu *et al.*, *Science* 261:209-211 (1993), which are herein incorporated by reference. Liposomes transfer genes to the target cells by fusing with the plasma membrane. The entry process is relatively efficient, but once inside the cell, the liposome-DNA complex has no inherent mechanism to deliver the DNA to the nucleus. As such, the most of the lipid and DNA gets shunted to cytoplasmic waste systems and destroyed. The obvious advantage of liposomes as a gene therapy vector is that liposomes contain no proteins, which thus minimizes the potential of host immune responses.

As another example, viral vector-mediated gene transfer is also a suitable method for the introduction of the vector into a target cell. Appropriate viral vectors include adenovirus vectors and adeno-associated virus vectors, retrovirus vectors and herpesvirus vectors.

Adenoviruses are linear, double stranded DNA viruses complexed with core proteins and surrounded by capsid proteins. The common serotypes 2 and 5, which are not associated with any human malignancies, are typically the base vectors. By deleting parts of the virus genome and inserting the desired gene under the control of a constitutive viral promoter, the virus becomes a replication deficient vector capable of transferring the exogenous DNA to differentiated, non-proliferating cells. To enter cells, the adenovirus fibre interacts with specific receptors on the cell surface, and the adenovirus surface proteins interact with the cell surface integrins. The virus penton-cell integrin interaction provides the signal that brings the exogenous gene-containing virus into a cytoplasmic endosome. The adenovirus breaks out of the endosome and moves to the nucleus, the viral capsid falls apart, and the exogenous DNA enters the cell nucleus where it functions, in an epichromosomal fashion, to express the exogenous gene. Detailed discussions of the use of adenoviral vectors for gene therapy can be found in Berkner, *Biotechniques* 6:616-629 (1988) and Trapnell, *Advanced Drug Delivery Rev.* 12:185-199 (1993), which are herein incorporated by reference. Adenovirus-derived vectors, particularly non-replicative adenovirus vectors, are characterized by their ability to accommodate exogenous DNA of 7.5 kB, relative stability, wide host range, low pathogenicity in man, and high titers (10^4 to 10^5 plaque forming units per cell). See Stratford-Perricaudet *et al.*, *PNAS* 89:2581 (1992).

Adeno-associated virus (AAV) vectors also can be used for the present invention. AAV is a linear single-stranded DNA parvovirus that is endogenous to many mammalian species. AAV has a broad host range despite the limitation that AAV is a defective parvovirus which is dependent totally on either adenovirus or herpesvirus for its reproduction *in vivo*. The use of AAV as a vector for the introduction into target cells of exogenous DNA is well-known in the art. *See, e.g., Lebkowski et al., Mole. & Cell. Biol.* 8:3988 (1988), which is incorporated herein by reference. In these vectors, the capsid gene of AAV is replaced by a desired DNA fragment, and transcomplementation of the deleted capsid function is used to create a recombinant virus stock. Upon infection the recombinant virus uncoats in the nucleus and integrates into the host genome.

Another suitable virus-based gene delivery mechanism is retroviral vector-mediated gene transfer. In general, retroviral vectors are well-known in the art. *See Breakfield et al., Mole. Neuro. Biol.* 1:339 (1987) and Shih *et al.*, in *Vaccines* 85: 177 (Cold Spring Harbor Press 1985). A variety of retroviral vectors and retroviral vector-producing cell lines can be used for the present invention. Appropriate retroviral vectors include Moloney Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus. These vectors include replication-competent and replication-defective retroviral vectors. In addition, amphotropic and xenotropic retroviral vectors can be used. In carrying out the invention, retroviral vectors can be introduced to a tumor directly or in the form of free retroviral vector producing-cell lines. Suitable producer cells include fibroblasts, neurons, glial cells, keratinocytes, hepatocytes, connective tissue cells, ependymal cells, chromaffin cells. *See Wolff et al., PNAS* 84:3344 (1989).

Retroviral vectors generally are constructed such that the majority of its structural genes are deleted or replaced by exogenous DNA of interest, and such that the likelihood is reduced that viral proteins will be expressed. *See Bender et al., J. Virol.* 61:1639 (1987) and Armento *et al., J. Virol.* 61:1647 (1987), which are herein incorporated by reference. To facilitate expression of the antisense or ribozyme molecule, of the inventive protein, a retroviral vector employed in the present invention must integrate into the genome of the host cell genome, an event which occurs only in mitotically active cells. The necessity for host cell replication effectively limits retroviral gene expression to tumor cells, which are highly

replicative, and to a few normal tissues. The normal tissue cells theoretically most likely to be transduced by a retroviral vector, therefore, are the endothelial cells that line the blood vessels that supply blood to the tumor. In addition, it is also possible that a retroviral vector would integrate into white blood cells both in the tumor or in the blood circulating through the tumor.

The spread of retroviral vector to normal tissues, however, is limited. The local administration to a tumor of a retroviral vector or retroviral vector producing cells will restrict vector propagation to the local region of the tumor, minimizing transduction, integration, expression and subsequent cytotoxic effect on surrounding cells that are mitotically active.

Both replicatively deficient and replicatively competent retroviral vectors can be used in the invention, subject to their respective advantages and disadvantages. For instance, for tumors that have spread regionally, such as lung cancers, the direct injection of cell lines that produce replication-deficient vectors may not deliver the vector to a large enough area to completely eradicate the tumor, since the vector will be released only from the original producer cells and their progeny, and diffusion is limited. Similar constraints apply to the application of replication deficient vectors to tumors that grow slowly, such as human breast cancers which typically have doubling times of 30 days versus the 24 hours common among human gliomas. The much shortened survival-time of the producer cells, probably no more than 7-14 days in the absence of immunosuppression, limits to only a portion of their replicative cycle the exposure of the tumor cells to the retroviral vector.

The use of replication-defective retroviruses for treating tumors requires producer cells and is limited because each replication-defective retrovirus particle can enter only a single cell and cannot productively infect others thereafter. Because these replication-defective retroviruses cannot spread to other tumor cells, they would be unable to completely penetrate a deep, multilayered tumor *in vivo*. See Markert *et al.*, *Neurosurg.* 77: 590 (1992). The injection of replication-competent retroviral vector particles or a cell line that produces a replication-competent retroviral vector virus may prove to be a more effective therapeutic because a replication competent retroviral vector will establish a productive infection that will transduce cells as long as it persists. Moreover, replicatively competent retroviral vectors may follow the tumor as it metastasizes, carried along and propagated by transduced tumor cells. The risks for complications are greater, with replicatively competent vectors, however.

Such vectors may pose a greater risk than replicatively deficient vectors of transducing normal tissues, for instance. The risks of undesired vector propagation for each type of cancer and affected body area can be weighed against the advantages in the situation of replicatively competent versus replicatively deficient retroviral vector to determine an optimum treatment.

Both amphotropic and xenotropic retroviral vectors may be used in the invention. Amphotropic viruses have a very broad host range that includes most or all mammalian cells, as is well known to the art. Xenotropic viruses can infect all mammalian cell cells except mouse cells. Thus, amphotropic and xenotropic retroviruses from many species, including cows, sheep, pigs, dogs, cats, rats, and mice, *inter alia* can be used to provide retroviral vectors in accordance with the invention, provided the vectors can transfer genes into proliferating human cells *in vivo*.

Clinical trials employing retroviral vector therapy treatment of cancer have been approved in the United States. See Culver, *Clin. Chem.* 40: 510 (1994). Retroviral vector-containing cells have been implanted into brain tumors growing in human patients. See Oldfield *et al.*, *Hum. Gene Ther.* 4: 39 (1993). These retroviral vectors carried the HSV-1 thymidine kinase (HSV-tk) gene into the surrounding brain tumor cells, which conferred sensitivity of the tumor cells to the antiviral drug ganciclovir. Some of the limitations of current retroviral based cancer therapy, as described by Oldfield are: (1) the low titer of virus produced, (2) virus spread is limited to the region surrounding the producer cell implant, (3) possible immune response to the producer cell line, (4) possible insertional mutagenesis and transformation of retroviral infected cells, (5) only a single treatment regimen of pro-drug, ganciclovir, is possible because the "suicide" product kills retrovirally infected cells and producer cells and (6) the bystander effect is limited to cells in direct contact with retrovirally transformed cells. See Bi *et al.*, *Human Gene Therapy* 4: 725 (1993).

Yet another suitable virus-based gene delivery mechanism is herpesvirus vector-mediated gene transfer. While much less is known about the use of herpesvirus vectors, replication-competent HSV-1 viral vectors have been described in the context of antitumor therapy. See Martuza *et al.*, *Science* 252: 854 (1991), which is incorporated herein by reference.

DIAGNOSTIC METHODS

The present invention also contemplates, for certain molecules described below, methods for diagnosis of human disease. In particular, patients can be screened for the occurrence of cancers, or likelihood of occurrence of cancers, associated with mutations in the encoded protein. DNA from tumor tissue obtained from patients suffering from cancer can be isolated and the gene encoding the protein can be sequenced. By examining a number of patients in this manner, mutations in the gene that are associated with a malignant cellular phenotype can be identified. In addition, correlation of the nature of the observed mutations with subsequent observed clinical outcomes allows development of prognostic model for the predicted outcome in a particular patient.

Screening for mutations conveniently can be carried out at the DNA level by use of PCR, although the skilled artisan will be aware that many other well known methods are available for the screening. PCR primers can be selected that flank known mutation sites, and the PCR products can be sequenced to detect the occurrence of the mutation. Alternatively, the 3' residue of one PCR primer can be selected to be a match only for the residue found in the unmutated gene. If the gene is mutated, there will be a mismatch at the 3' end of the primer, and primer extension cannot occur, and no PCR product will be obtained. Alternatively, primer mixtures can be used where the 3' residue of one primer is any nucleotide other than the nonmutated residue. Observation of a PCR product then indicates that a mutation has occurred. Other methods of using, for example, oligonucleotide probes to screen for mutations are described, for example, in U.S. Patent No. 4,871,838, which is herein incorporated by reference in its entirety.

Alternatively, antibodies can be generated that selectively bind either mutated or non-mutated protein. The antibodies then can be used to screen tissue samples for occurrence of mutations in a manner analogous to the DNA-based methods described *supra*.

The diagnostic methods described above can be used not only for diagnosis and for prognosis of existing disease, but may also be used to predict the likelihood of the future occurrence of disease. For example, clinically healthy patients can be screened for mutations in the inventive molecule that correlate with later disease onset. Such mutations may be observed in the heterozygous state in healthy individuals. In such cases a single mutation event can effectively disable proper functioning of the gene and induce a transformed or malignant phenotype. This screening also may be carried out prenatally or neonatally.

DNA molecules according to the invention also are well suited for use in so-called "gene chip" diagnostic applications. Such applications have been developed by, *inter alia*, Synteni and Affymetrix. Briefly, all or part of the DNA molecules of the invention can be used either as a probe to screen a polynucleotide array on a "gene chip," or they may be immobilized on the chip itself and used to identify other polynucleotides via hybridization to the surface of the chip. In this manner, for example, related genes can be identified, or expression patterns of the gene in various tissues can be simultaneously studied. Such gene chips have particular application for diagnosis of disease, or in forensic analysis to detect the presence or absence of an analyte. Suitable chip technology is described for example, in Wodicka *et al.*, *Nature Biotechnology*, 15:1359 (1997) which is hereby incorporated by reference in its entirety, and references cited therein.

PROTEIN-PROTEIN INTERACTIONS

Due to their similarity to certain known proteins, it is anticipated that some of the inventive protein molecules will interact with another class of cellular proteins. This is particularly true of those molecule containing leucine zipper motifs.

Any method suitable for detecting protein-protein interactions can be employed for identifying interacting targets. Among the traditional methods which can be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Utilizing procedures such as these allows for the identification of GAP gene products. Once identified, a GAP protein can be used, in conjunction with standard techniques, to identify its corresponding pathway gene. For example, at least a portion of the amino acid sequence of the pathway gene product can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique (see, *e.g.*, Creighton, 1983, *PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES*, W.H. Freeman & Co., N.Y., pp.34-49). The amino acid sequence obtained can be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for pathway gene sequences. Screening can be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and for screening are well-known. (See *e.g.*, Ausubel, *supra*, and *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, 1990, Innis *et al.*, eds. Academic Press, Inc., New York).

Additionally, methods can be employed which result in the simultaneous identification of interacting target genes. One method which detects protein interactions *in vivo*, the two-hybrid system, is described in detail for illustration purposes only and not by way of limitation. One version of this system has been described (Chien *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 9578-9582 (1991)) and is commercially available from Clontech (Palo Alto, CA).

Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to a known protein, in this case an inventive protein, and the other contains the activator protein's activation domain fused to an unknown protein (a putative GAP, for instance) that is encoded by a cDNA which has been recombined into this plasmid as part of a cDNA library. The plasmids are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (*e.g.*, *lacZ*) whose regulatory region contains the transcription activator's binding sites. Either hybrid protein alone cannot activate transcription of the reporter gene, the DNA-binding domain hybrid cannot because it does not provide activation function, and the activation domain hybrid cannot because it cannot localize to the activator's binding sites. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

The two-hybrid system or related methodology can be used to screen activation domain libraries for proteins that interact with a known "bait" gene product. By way of example, and not by way of limitation, gene products known to be involved in TH cell subpopulation-related disorders and/or differentiation, maintenance, and/or effector function of the subpopulations can be used as the bait gene products. Total genomic or cDNA sequences are fused to the DNA encoding on activation domain. This library and a plasmid encoding a hybrid of the bait gene product fused to the DNA-binding domain are cotransformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene. For example, and not by way of limitation, the bait gene can be cloned into a vector such that it is translationally fused to the DNA encoding the DNA-binding domain of the GAL4 protein. These colonies are purified and the library plasmids responsible for reporter gene expression are isolated. DNA sequencing is then used to identify the proteins encoded by the library plasmids.

The present invention, thus generally described, will be understood more readily by reference to the following examples, which are provided by way of illustration and are not intended to be limiting of the present invention.

The examples below are provided to illustrate the subject invention. These examples are provided by way of illustration and are not included for the purpose of limiting the invention.

EXAMPLES

EXAMPLE I: cDNA Library Construction

cDNA library plates and clones originated from five cDNA libraries that were constructed by directional cloning. These are available through the Resource Center (<http://www.rzpd.de>) of the German Genome Project. In particular, the hfbr2 (human fetal brain; RZPD number DKFZp564) and hfkd2 (human fetal kidney; DKFZp566) libraries were generated using the Smart kit (Clontech), except that PCR was carried out with primers that contained uracil residues to permit directional cloning without restriction digestion and ligation, and were complementary with the pAMP1 (LifeTechnologies) cloning sites for directional cloning. The htes3 (human testes; DKFZp434), hute1 (human uterus; DKFZp586) and hmcfl (human mammary carcinoma; DKFZp727) libraries are conventional (Gubler, U., Hoffman, B.J., (1983), A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269), size-selected cDNA libraries. They are cloned into pSPORT1 (LifeTechnologies) via a NotI site which is introduced during reverse transcription downstream of the oligo dT primer and a SalI site that is introduced by the ligation of a adapters. The human mammary carcinoma library was constructed from MCF7 cells.

The cDNA sequences of this application were first identified among the sequences comprising various libraries. Technology has advanced considerably since the first cDNA libraries were made. Many small variations in both chemicals and machinery have been instituted over time, and these have improved both the efficiency and safety of the process. Although the cDNAs could be obtained using an older procedure, the procedure presented in this application is exemplary of one currently being used by persons skilled in the art. For the

purpose of providing an exemplary method, the mRNA isolation and cDNA library construction described here is for the MCF-7 library (DKFZp727) from which the clones named DKFZphmcf1_xxyyxx were obtained.

The human cell line MCF-7 was grown in DMEM supplemented with 10% fetal calf serum until confluency. 3×10^8 cells were harvested with a cell scraper in PBS. Cells were lysed in buffer containing 0.5 % NP-40 to leave the nuclei intact. The debris was pelleted by centrifugation at $15\,000 \times g$ for 10 minutes at 4 degrees Celsius. Proteins in the supernatant were degraded in presence of SDS and Proteinase K (30 minutes at 56 degrees Celsius). Precipitation of proteins was done in a Phenol/Chloroform extraction, RNA was precipitated from the aqueous phase with Na-acetate and Ethanol. Polyadenylated messages were isolated using Qiagen Oligotex (QIAGEN, Hilden Germany).

First strand cDNA synthesis was accomplished using an oligo (dT) primer which also contained an NotI restriction site. Second strand synthesis was performed using a combination of DNA polymerase I, *E. coli* ligase and RNase H, followed by the addition of a Sall adaptor to the blunt ended cDNA. The Sall adapted, double-stranded cDNA was then digested with NotI restriction enzyme, and fractionated by size on an agarose gel. DNA of the appropriate size was cut from the gel and cast into a second gel in a 90° angle. After electrophoresis in the second dimension, cDNA of the appropriate size was cut from the gel. The agarose block was broken down with help of gelase. The cDNA was purified with help of two phenol extractions and an ethanol precipitation. The cDNA was ligated into Sall/NotI pre-digested pSport1 vector (LifeTechnologies) and transformed into DH10B bacteria.

The libraries were arrayed into 384-well microtiter plates and spotted on high density nylon membranes for hybridization analysis. Filters and clones are available through the Resource Center. Whole plates were distributed to the sequencing partners of the consortium for systematic sequencing.

EXAMPLE II: Sequencing of cDNA Clones

All clones in the 384-well microtiter plates were sequenced from the 5' end. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on

ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry.

The resulting expressed sequence tag (EST) sequences ("r1 ESTs" = sequenced from 5'-end) were analysed for:

a) the lack of identical matches with known genes.

For this, the EST-sequence was blasted against the cDNA consortiums own database and after that against public databases and (with BLASTn and BLASTx against EMBL/EMBLNEW and assembled ESTs, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings). ESTs which were identical to known genes in more than 100 bp, with less than 2 mismatches, were excluded from further analysis.

b) the presence of an open reading frame

Open reading frames (ORFs) were detected with a tool developed by Munich Information Center for Protein Sequences (MIPS) called ORF-map. ORF-map visualises potential start and stop-codons. If an ORF without a stop codon was detected in a r1-EST, the sequence was processed further.

c) the presence of GC rich sequences

A script developed by MIPS computed the GC-content of the r1-sequence, which should be >40%. Writing similar scripts is within the ordinary skill of one in bioinformatics.

d) the lack of repeat structures

Repeats such as Alu, Line or CA-repeats were detected by blasting (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings) against a repeat-database compiled by MIPS. If a repeat was present within the r1-sequence, the sequence were not processed further.

Novel clones that met all criteria were identified to the sequencers, who then performed 3'-end sequencing of these clones. The resulting 3' ESTs ("s1 ESTs" = sequenced from 3'-end) were checked for

a) the lack of matches with known genes in public databases, and sequences already generated by us.

This was done by blasting against EMBL/EMBLNEW and assembled EST (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings).

b) the presence of polyadenylation signals.

Again only clones matching the selection criteria were chosen to be sequenced completely by the sequencers. Clones were selected after the following criteria:

A very good ORF had at least one BLASTx match to other proteins. A "good ORF" should extend to the 3' end and be longer than ~40 codons. If the ORF started in the r1 sequence, in front of the potential start codon, there should not exist too many competing start codons in frame with the ORF start codon and the start should match the Kozak consensus ATG. If the EST sequence was too short to decide according to the potential ORF, and there were only a few or no start codons in the sequence the GC content of the Sequence should be greater than 40%. The r1 sequences needed not contain a polyA-tail at the 3' end. In addition, the results of the blasting against the assembled human ESTs could help in questionable cases to decide whether to stop or to continue. A hit against these ESTs was an indication to go further.

Clones passing the above-described screening were sequenced in full. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry. Primer walking (Strauss et al., 1986, Specific-primer-directed DNA sequencing. *Anal Biochem.* 154, 353-360) was the preferred sequencing strategy because of the lower redundancy possible compared to random shotgun (Messing, J., Crea, R., Seeburg, H.P. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9, 32-39) methods. Walking primers were generally designed using software (e.g. Haas, S., Vingron, M., Poustka, A., Wiemann, S. (1998) Primer design in large-scale sequencing. *Nucleic Acids Res.* 26, 3006-3012, Schwager, C., Wiemann, S., Ansorge, W. (1995) GeneSkipper: integrated software environment for DNA sequence assembly and

alignment. HUGO Genome Digest 2, 8-9) that permitted complete automation of this usually time consuming process and helped in the parallel processing of large numbers of clones.

EXAMPLE III: Bioinformatics analysis of full length cDNAs

Each sequence obtained was compared on nucleotide level in a stepwise manner to sequences in EMBL/EMBLNEW, EMBL-EST, EMBL-STS using the BLASTn algorithm. Basic Local Alignment Search Tool (BLAST, Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S. F. et al (1990) J Mol Biol 215:403-10) is used to search for local sequence alignments. BLAST produces alignments of both nucleotide (BLASTn) and amino acid sequences (BLASTp or BLASTx) to determine sequence similarity. BLAST is especially useful in determining exact matches or in identifying homologs, because of the local nature of the alignments. While it is useful for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment BLAST approach is to look threshold or cut off score set by the user. BLAST looks for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output. Parameter settings for the BLAST-operations (BLASTN 2.0a19MP-WashU) described were: EMBL-EMBLNEW: H=0 V=5 B=5 -filter seg; EMBL-EST: H=0 E=1e-10 B=500 V=500 -filter seg; EMBL-STS: H=0 V=5 B=5.

Search against EMBL/EMBLNEW was done to determine whether the cDNAs are already known, and also to find out whether the cDNAs are encoded by genomic sequences already sequenced and published/submitted to these databases.

Search against EMBL-EST was performed to get a first impression how abundant a particular cDNA would be and to get information on tissue specificity (so-called “electronic Northern-Blot”, e.g. some of the cDNAs derived of the testis library show only hits to ESTs also derived of testis libraries).

The cDNA-sequences were blasted against EMBL-STS to determine STS-sequence-match to the cDNA, thus providing a mapping information to the new cDNA.

The potential protein-sequences were generated automatically by a script searching for the longest open reading frame (ORF) in each of the three forward frames with a minimum length of 90 codons. Next, the automatically generated ORFs were translated into protein sequences. These protein sequences were searched against the non redundant protein data set of PIR/SwissProt/Trembel/Tremblnew (BLASTP 2.0a19MP-WashU, parameter setting: V=7 B=7 H=0 -filter seg). If the script generated more than one ORF, one ORF was chosen manually by the annotater according to the degree of similarity to known proteins, the location of the ORF in the cDNA, the length, the amino acid composition and the content of Prosite-Motifs.

Additionally there was a BLASTx (BLASTX 2.0a19MP-WashU against non redundant protein database comprising PIR/SWISSPROT/TREMBL/TREMBLNEW; parameter-settings were: matrix/home/data/blast/matrix/aa/BLOSUM62 H=0 V=5 B=5 -filter seg) search to find potential frame shift in the complementary cds of the cDNAs and to identify unspliced or partly spliced cDNAs. The protein sequence was then transferred to the PEDANT system, in order to generate additional information on the new proteins. PEDANT (Protein Extraction, Description, and ANalysis Tool, Frishman, D. & Mewes, H.-W. (1997) PEDANTic genome analysis. Trends in Genetics , 13, 415-416) is a platform developed at the Munich Information Center for Protein Sequences (MIPS, Munich, Germany), which incorporates practically all bioinformatics methods important for the functional and structural characterisation of protein sequences. Computational methods used by PEDANT are:

FASTA

Very sensitive protein sequence database searches with estimates of statistical significance. Pearson W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* 183, 63-98.

BLAST2

Very sensitive protein sequence database searches with estimates of statistical significance. Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-10.

PREDATOR

High-accuracy secondary structure prediction from single and multiple sequences. Frishman, D. and Argos, P. (1997) 75% accuracy in protein secondary structure prediction. *Proteins*, 27, 329-335. Frishman, D. and Argos, P. (1996) Incorporation of long-distance interactions in a secondary structure prediction algorithm. *Prot. Eng.* 9, 133-142.

STRIDE

Secondary structure assignment from atomic coordinates. Frishman, D. and Argos, P. (1995) Knowledge-based secondary structure assignment. *Proteins* 23, 566-579.

CLUSTALW

Multiple sequence alignment. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680.

TMAP

Transmembrane region prediction from multiply aligned sequences. Persson, B. and Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.* 237, 182-192.

ALOM2

Transmembrane region prediction from single sequences. Klein, P., Kanehisa, M., and DeLisi, C. Prediction of protein function from sequence properties: A discriminant analysis of a database. *Biochim. Biophys. Acta* 787, 221-226 (1984). Version 2 by Dr. K. Nakai.

SIGNALP

Signal peptide prediction Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* 10, 1-6.

SEG

Detection of low complexity regions in protein sequences. Wootton, J.C., Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Computers & Chemistry* 17, 149-163.

COILS

Detection of coiled coils. Lupas, A., M. Van Dyke, and J. Stock, "Predicting Coiled Coils from Protein Sequences." *Science* (1991) 252, 1162-1164.

PROSEARCH

Detection of PROSITE protein sequence patterns. Kolakowski L.F. Jr., Leunissen J.A.M., Smith J.E. (1992) ProSearch: fast searching of protein sequences with regular expression patterns related to protein structure and function. *Biotechniques* 13, 919-921.

BLIMPS

Similarity searches against a database of ungapped blocks. J.C. Wallace and Henikoff S., (1992) PATMAT: a searching and extraction program for sequence, pattern and block queries and databases, *CABIOS* 8, 249-254. Written by Bill Alford.

HMMER

Hidden Markov model software . Sonnhammer E.L.L., Eddy S.R., Durbin R. (1997)
Pfam: A Comprehensive Database of Protein Families Based on Seed Alignments. *Proteins*
28, 405-420.

pI

Perl script that returns the amino acid composition, molecular weight, theoretical pI, and expected extinction coefficient of an amino acid sequence. By Fred Lindberg. The parameter-settings were as follows: known3d: score > 100; BLAST: E-value < 10; SCOP: <= 50 Alignments, E-Value < 0.0001; signalp: Y=0.7; untersucht vom N-Terminus her: 50 aa; funcat: E-value < 0.001; BLOCKS: <= 10 hits; BLIMPS: threshold 1100.0; COILS: threshold 0.95; SEG: threshold 20.0; BLAST in report: E-value < 0.001; PIR-KW, superfamilies, EC-Nummern in report: E-value < 0.00001; known3d in report: score > 120

The results of PEDANT analysis, together with the results of the similarity searches, constitute the basis for the structural and functional annotation of the cDNAs and the encoded proteins, as specified below.

EXAMPLE III: CELLULAR LOCALIZATIONS OF GFP-FUSION PROTEINS

Plasmids of cDNA-GFP fusions were transfected into mammalian tissue culture cells and allowed to express the proteins for up to 48 hours. Live cells were imaged at 24 hours and 48 hours after transfection and the localisations recorded. The chart, below, depicts the apparent final cellular localisations of 107 cDNA-GFP fusions.

In order to minimize the possibility of the GFP interfering with protein function and/or localization, two separate populations of cDNAs were generated encoding N-terminal or C-terminal GFP fusions. Clearly this appears to be a crucial strategy, since overall only 56% of the proteins localised to a specific compartment irrespective of the position of the GFP. In the instances where only one fusion localized, the complementary fusion either gave no expression or a nuclear and cytosolic staining - characteristic for GFP alone expression.

Each cDNA in turn was subjected to bioinformatic analysis. Where possible, the potential subcellular localisations of the expressed proteins were determined. This

information was then compared to the actual localisations determined from expression of the GFP-fusion proteins in mammalian cells.

DKFZphfbr2_16c16

group: Cell structure and motility

DKFZphfbr2_16c16.3 encodes a novel 586 amino acid protein with similarity to the human actin binding protein MAYVEN and Drosophila Kelch.

MAVEN is a novel actin binding protein predominantly expressed in brain. Drosophila kelch is involved in the maintenance of ring canal organization during oogenesis. The amino half of the protein including the BTB domain mediates dimerization, while the amino half might allow cross-linking of ring canal actin filaments, thus organising the inner rim cytoskeleton. The kelch repeat domain is necessary for ring canal localisation and believed to mediate an additional interaction, possibly with actin. The new protein shares the features of both proteins and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins.

The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

similarity to Drosophila kelch

complete cDNA, complete cds, EST hits
on genomic level partly encoded by AC005082 and AC006039

Sequenced by Qiagen

Locus: unknown

Insert length: 3028 bp

Poly A stretch at pos. 3004, polyadenylation signal at pos. 2984

```
1 GGGGGCCCGG GGACGCAGCC CAGTTGGTAG CGTCGCTCCC TGAGCGTTTC
51 TAAGGGGGCC GCCCGGCCCT GTCTTTGGGC AGTGGCCGAG CCACCGCCCG
101 CTGCCGCGCG TTCCAGAGCT GGGCGCTGCA GCTGCACTGC CGATCGCCGT
151 GTTTGGTCCG TAGAATCCCC AGTGTGCCCA GAGAGTCCGA CCCCTCGCCC
201 GCGCCGCGCG GCCCGGGCG TGAACCGAGC TGAGGGAGGA TGGCAGCCTC
251 TGGGGTGGAG AAGAGCAGCA AGAAGAAGAC CGAGAAGAAA CTTGCTGCTC
301 GGAAGAAGC TAAATTGTTG GCGGGTTTCA TGGGCGTCAT GAATAACATG
351 CGGAACAGCA AAACGTTGTG TGACGTGATC CTCATGGTCC AGGAAAGAAA
401 GATACCTGCT CATCGTGTG TTCTTGCTGC AGCCAGTCAT TTTTAACT
451 TAATGTTTAC AACTAACATG CTTGAATCAA AGTCCTTTGA AGTAGAACTC
501 AAAGATGCTG AACCTGATAT TATTGAACAA CTGGTGGAAAT TTGCTTATAC
551 TGCTAGAATT TCCGTGAATA GCAACAATGT TCAGTCTTTG TTGGATGCAG
601 CAAACCAATA TCAGATTGAA CCTGTGAAGA AAATGTGTGT TGATTTTTTG
651 AAAGAACAAG TTGATGCTTC AAATTGTCTT GGTATAAGTG TGCTAGCGGA
701 GTGTCTAGAT TGTCTGAAT TGAAGCAAC TGCAGATGAC TTTATTCATC
751 AGCACTTTAC TGAAGTTTAC AAAACTGATG AATTTCTTCA ACTTGATGTC
801 AAGCGAGTAA CACATCTTCT CAACCAGGAC ACTCTGACTG TGAGAGCAGA
851 GGATCAGGTT TATGATGCTG CAGTCAGGTG GTTGAATAAC GATGAGCCTA
901 ATCGCCAGCC ATTTATGGTT GATATCCTTG CTAAGTCAG GTTTCCTCTT
951 ATATCAAAGA ATTTCTTAAG TAAAACGGTA CAAGCTGAAC CACTTATTC
1001 AGACAATCCT GAATGCCTTA AGATGGTGAT AAGTGAATG AGGTACCATC
1051 TACTGTCTCC AGAGGACCGA GAAGAACTTG TAGATGGCAC AAGACCTAGA
1101 AGAAAGAAAC ATGACTACCG CATAGCCCTA TTTGGAGGCT CTCAACCACA
1151 GTCTTGTAGA TATTTTAACC CAAAGGATTA TAGCTGGACA GACATCCGCT
1201 GCGGCTTTGA AAAACGAAGA GATGCAGCAT GCGTGTTTTG GGACAATGTA
1251 GTATACATTT TGGGAGGCTC TCAGCTTTTC CCAATAAAGC GAATGGACTG
1301 CTATAATGTA GTGAAGGATA GCTGGTATTC GAAACTGGGT CCTCCGACAC
1351 CTCGAGACAG CCTTGCTGCA TGTGCTGCAG AAGGCAAAAT TTATACATCT
1401 GGAAGGTCAG AAGTAGGAAA CTCAGCTCTG TATTTATTTG AGTGCTATGA
1451 TAGAGAAGCT GAAAGCTGGC ACACAAGGCC CAGCATGCTG ACCCAGCGCT
1501 GCAGCCATGG GATGGTGGA GCCAATGGCC TAATCTATGT TTGTGGTGGA
1551 AGTTTAGGAA ACAATGTTTC AGGGAGAGTG CTTAATTCCT GTGAAGTTTA
1601 TGATCCTGCC ACAGAAACAT GGACTGAGCT GTGTCCAATG ATTGAAGCCA
1651 GGAAGAATCA TGGGCTGGTA TTTGTAAAAG ACAAGATATT TGCTGTGGGT
1701 GGTGAGAATG GTTTAGGTGG TCTGGACAA GTGGAATATT ACGATATTAA
1751 GTTGAACGAA TGGAAGATGG TCTACCAAT GCCATGGAAG GGTGTAACAG
1801 TGAAATGTGC AGCAGTTGGC TCTATAGTTT ATGTCTTGCC TGGTTTTTCAG
```

```

2301 AGAAGATTGG CTCATCAGTG AAGCGCAGTA TCTTAGCTCT AGATTCTATT
2351 TTCATGCATC ACAGAAAGTGC TATACGGTTA GGTCTGTTTG TGCTCAGTCA
2401 AGAACTAAGA AATAGTATGA ATGTGAAGTC AAGATGGGCA ACTCAGATGG
2451 AGCAGCTTAG TCTCAGATT TGCTTGCTA TTTATTTTAT TTAGTGCCAA
2501 ATGTATTCCA TTTTAAAAGT AAGCCAGAGT GAGTCAAGGC ATATACACAC
2551 TTTCTCACAA AACTTCCTAA ACAGATTGG GGGTTTAATA TGCCAACTC
2601 CTCATGAAAT ATATTCAATC CACTTAAATA TATTCCATCT TTTTAAACATA
2651 AAATGTAAAG CTTAGCACCC ATCATTAAAT TATGTCCTCTG TTTTATCCAG
2701 TGGTTAAAAA AGGATTCTGC CTCTTTAGTC CTCACGTGTA AATAAAACCC
2751 AATCATAGTA AGTGATTAACT TAGCAAAAAG TAAAGCTATT TATAGCAAAT
2801 TTCTAGATCA TTAGAAAAGC ACTGGTAGTT GTACAATATC AGTGTTGACT
2851 TTGAACCTTCT TTAACGAGAT CATGAATTCT TTCCCTTAG CCAAAACATG
2901 AAATATTTAA CCTAGTTGTC TCTAAAAGTT TTGTAATCAT GAGTTAGATA
2951 TATGTCATCT CCTATTTCATT GCTTTTATGT GATCAATAAA TCTTTTACAA
3001 ACCCAAAAAGA AAAAAAAAAA AAAAAAAA

```

BLAST Results

Entry AC005082 from database EMBL:
Homo sapiens clone RG271G13; HTGS phase 1, 7 unordered pieces.
Score = 6460, P = 0.0e+00, identities = 1292/1292
4 exons matching Bp 1180-3007

Entry AC006039 from database EMBL:
*** SEQUENCING IN PROGRESS *** Homo sapiens clone NH0319F03; HTGS phase 1, 3 unordered pieces.
Score = 1780, P = 2.0e-117, identities = 368/377
5 exons matching Bp 6-860

Entry HSG20603 from database EMBL:
human STS A005Y34.
Score = 670, P = 1.0e-23, identities = 134/134

Medline entries

93201592:
kelch encodes a component of intercellular bridges in
Drosophila egg chambers.

97412177:
Drosophila kelch is an oligomeric ring canal actin organizer.

Peptide information for frame 3

ORF from 240 bp to 1997 bp; peptide length: 586
Category: strong similarity to known protein

```

1 MAASGVKSS KKKTEKKLAA REEAKLLAGF MGVMMNMRKQ KTLCDVILMV
51 QERKIPAHRV VLAAASHFFN LMFTTNMLES KSFEVELKDA EPDIIEQLVE
101 FAYTARISVN SNNVQSLLDA ANQYQIEPVK KMCVDFLKEQ VDASNCLGIS
151 VLAELCDCEP LKATADDFIH QHTEVYKTD EFLQLDVKRV THLLNQDTLT
201 VRAEDQVYDA AVRWLKYDEP NRQPEMVDIL AKVRFPLISK NFLSKTVQAE
251 PLIQDNPECL KMVISGMRYH LLSPEDREEL VDGTRPRRKK HDYRIALFGG
301 SQPQSCRYFN PKDYSWTDIR CPFEEKRRDAA CVFWDNVVYI LGGSQLFPIK
351 RMDCYNVVKD SWYSKLGPPPT PRDSLAAACAA EGKIYTSGGG EVGNSALYLF
401 ECDYTRTESW HTKPSMLTQR CSHGMVEANG LIYVCGGSLG NNVSGRVLNS
451 CEVYDPATET WTELCPMIEA RKNHGLVFKV DKIFAVGGQN GLGGLDNVEY
501 YDIKLNEWKM VSPMPWKGVT VKCAAVGSIV YVLAFQGVG RLGHILEYNT
551 ETDKVVANSK VRAFPVTSCS ICVVDTCGAN EETLET

```

BLASTP hits

Entry KELC_DROME from database SWISSPROT:
RING CANAL PROTEIN (KELCH PROTEIN).
Length = 689
Score = 816 (287.2 bits), Expect = 1.9e-81, P = 1.9e-81
Identities = 187/542 (34%), Positives = 290/542 (53%)

Entry AC004021.1 from database TREMBL:
WUGSC:H_DJ0186K10.1"; Human PAC clone DJ0186K10 from 5q31,
complete sequence. Homo sapiens (human)
Length = 497

Entry A45773 from database PIR:
kelch protein, long form - fruit fly (*Drosophila melanogaster*)
Length = 1476
Score = 817 (287.6 bits), Expect = 1.7e-80, P = 1.7e-80
Identities = 189/549 (34%), Positives = 292/549 (53%)

Report for DKFZphfbr2 16c16.3

```

[BLOCKS]      BL00075D Dihydrofolate reductase proteins
[SCOP]        dlqog_3 2.46.1.1.1 (151-537) Galactose oxidase, central domain 6e-36
[PIRKW]       zinc finger 2e-11
[PIRKW]       DNA binding 9e-10
[PIRKW]       transcription factor 1e-06
[SUPFAM]      A55R protein middle region homology 1e-35
[SUPFAM]      POZ domain homology 1e-35
[SUPFAM]      vaccinia virus 59K HindIII-C protein 5e-15
[SUPFAM]      A55R protein 1e-35
[SUPFAM]      myxoma virus M9-R protein 2e-11
[SUPFAM]      A55R protein carboxyl-terminal homology 1e-35
[PROSITE]     CAMP_PHOSPHO_SITE      2
[PROSITE]     MYRI_STYL      8
[PROSITE]     CK2_PHOSPHO_SITE      10
[PROSITE]     TYR_PHOSPHO_SITE      1
[PROSITE]     PKC_PHOSPHO_SITE      11
[PROSITE]     ASN_GLYCOSYLATION      1
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY      3.75 %

```

SEQ SWYSLGPPTPRDSLAAEAGKIYTSGGSEVGNALYLFECDTRTESWHTKPSMLTQR
 SEG
 PRD CC

```

SEQ      CSHGMVEANGLIYVCGGSLGNNVSGRVLNSCEVYDPATETWTELCPMIEARKNHGLVFK
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      DKIFAVGGQNLGLDNDVEYYDIKLNEWKMVSPMPWKGVTVKCAAVGSIVYVLAGFQGVG
SEG      .....
PRD      ceeeecccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      RLGHILEYNTETDKWVANSKVRAPVTSCLICVVDTCGANEETLET
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_16cl6.3

PS00001	442->446	ASN_GLYCOSYLATION	PDOC00001
PS00004	11->15	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	188->192	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	552->555	PKC_PHOSPHO_SITE	PDOC00005
PS00006	4->8	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	315->319	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	405->409	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	550->554	CK2_PHOSPHO_SITE	PDOC00006
PS00007	202->209	TYR_PHOSPHO_SITE	PDOC00007
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	389->395	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	436->442	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00008	493->499	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16cl6.3)

DKFZphfbr2_16f21

group: brain derived

DKFZphfbr2_16f21 encodes a novel 208 amino acid protein with strong similarity to human zinc finger protein 216.

The novel protein shows strong similarity to the human zinc finger protein 216, but has no Zn finger.

PROSITE: Contains no Zinc finger; No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to zinc finger protein 216

complete cDNA, complete cds, EST hits
start matches Kozak consensus ANNatgG,

Sequenced by Qiagen

Locus: unknown

Insert length: 1512 bp

Poly A stretch at pos. 1490, polyadenylation signal at pos. 1474

```
1 GGGAGCAAGC AGGGGTTTCGG CGGCATTACC TGTACCCATT CACCGGCGGC
51 TACCGGCGGC GCGCGTAGC GTGTCAGGCG GAGAGACCCG CCGCCAGGTG
101 TGCAACTGAG GAACATGGCT CAAGAACTA ATCACAGCCA AGTGCCTATG
151 CTTTGTTCCTA CTGGCTGTGG ATTTTATGGA AACCTCGTA CAAATGGCAT
201 GTGTTCACTA TGCTATAAAG AACATCTTCA AAGACAGAAT AGTAGTAATG
251 GTAGAATAAG CCCACCTGCA ACCTCTGTCA GTAGTCTGTC TGAATCTTTA
301 CCAGTTCAAT GCACAGATGG CAGTGTGCCA GAAGCCAGT CAGCATTAGA
351 CTCTACATCT TCATCTATGC AGCCAGCCCC TGTATCAAAT CAGTCACTTT
401 TATCAGAATC TGTAGCATCT TCTCAATTGG ACAGTACATC TGTGGACAAA
451 GCAGTACCTG AAACAGAAGA TGTGCAGGCT TCAGTATCAG ACACAGCACA
501 GCAGCCATCT GAAGAGCAAA GCAAGCCTCT TGA AAAACCGG AAACAAAAAA
551 AGAATCGCTG TTTCATGTGC AGGAAGAAAG TGGGACTTAC TGGGTTTGAA
601 TGCCGGTGTG GAAATGTTTA CTGTGGTGTA CACCGTTACT CAGATGTAAT
651 CAATTGCTCT TACAATTACA AAGCCGATGC TGCTGAGAAA ATCAGAAAAAG
701 AAAATCCAGT AGTTGTTGGT GAAAAGATCC AAAAGATTG AACTCCTGCT
751 GGAATACAAA ATTCTTGAGC ATCTGCAAA C TAAAAATGA CTTGAGGTTT
801 TTTTTCCTT AGTCATTGGG AATGTAGAGC AGTGTATCTT GCATGTCATC
851 GGAAGAATAG ATTTTGTGTT TGGTTTGTG TTGAAAATGA CTCTGAACAT
901 TTATTTCCTA TGCAATTCT GTGGCTGAGG AGACTTAAAC TTTACAAGTA
951 TTATCCTTTT AAGATCATTT TAATTTTAGT TGAGTGCAGA GGGCTTTTAT
1001 AACAAACGTG CAGAAATTTT GGAGGGCTGT GATTTTCCA GTATTAAACA
1051 TGCAATGATT AATCTTGCGA TTTATTTTCT CATTATGTAT GTATATATCG
1101 CTTTCTCTCG CAGCAGGATT TCTCTTTTGA TAATGCCCTT TAGGGCACAA
1151 CTAGTTATCA GTAACCTGAT GTATCTTAAT CATTATGGCT GCTTCTGTTT
1201 TTTTATTAAC AAAGGTTATT CATATGTTAG CATATAGTTT CTTTGACCCC
1251 ACTATTATG TCTGAATCAT TTGTCACAAG AGAGTGTGTG CTGATGAGAT
1301 TGTAAAGTTG TGTGTTTAAA CTTTTTTTGG AGCGAGGGAA GAAAAAGCTG
1351 TATGCATTTT ATTGCTGTCT ACAGGTTTCT TTCAGATTAT GTTCATGGGT
1401 TTGTGTGTAT ACAATATGAA GAATGATCTG AAGTAATTGT GCTGTATTTA
1451 TGTTTATTC CAGTCTTTG ATTAATAAAA AAGGAAAACC AGAAAAAAA
1501 AAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 115 bp to 738 bp; peptide length: 208
Category: strong similarity to known protein

```

1 MAQETNHSQV PMLCSTGCGF YGNPRTNGMC SVCYKEHLQR QNSSNGRISP
51 PATSVSSLSE SLPVQCTDGS VPEAQSLDS TSSSMQSPV SNQSLLESV
101 ASSQLDSTSV DKAVPETEDV QASVSDTAQQ PSEEQSKPLE KPKQKKNRCF
151 MCRKKVGLTG FECRCGNVYC GVHRYSDVLN CSYNYKADAA EKIRKENPVV
201 VGEKIQKI

```

BLASTP hits

Entry ATF7H19_1 from database TREMBLNEW:
gene: "F7H19.10"; product: "putative protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone F7H19 (ESSAII project) >TREMBL:ATT12H17_21
gene: "T12H17.210"; product: "predicted protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAII project)
Score = 206, P = 2.1e-24, identities = 51/146, positives = 77/146

Entry PVPVPR3A_1 from database TREMBL:
gene: "PVPVPR3"; P.vulgaris PVPVPR3 protein mRNA, complete cds.
Score = 237, P = 4.9e-20, identities = 50/136, positives = 73/136

Entry AF062072_1 from database TREMBL:
gene: "ZNF216"; product: "zinc finger protein 216"; Homo sapiens zinc finger protein 216 (ZNF216) gene, complete cds.
Score = 591, P = 1.6e-57, identities = 124/215, positives = 147/215

Alert BLASTP hits for DKFZphfbr2_16f21, frame 1

TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus zinc finger protein 216 mRNA, complete cds., N = 1, Score = 590, P = 2.1e-57

TREMBLNEW:AB001773_1 gene: "pem-6"; product: "PEM-6"; Ciona savignyi pem-6 (posterior end mark 6) mRNA, complete cds., N = 1, Score = 421, P = 1.7e-39

>TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus zinc finger protein 216 mRNA, complete cds.
Length = 213

HSPs:

Score = 590 (88.5 bits), Expect = 2.1e-57, P = 2.1e-57
Identities = 123/213 (57%), Positives = 146/213 (68%)

```

Query:      1 MAQETNHSQV PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPPAT---SVSS 57
             MAQETN + PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQ +S GR+SP T S S
Sbjct:      1 MAQETNQTG PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNS-GRMSPMGTASGSNSP 59

Query:      58 LSESLPVQCTDGSVPEAQSLDSTSSSMQSPVSNQSLLE--SVASSQLDSTSVDKAVP 115
             S+S VQ D + + A STS + PV+ + + ++ S+ D + K
Sbjct:      60 TSDSASVQRADAGLNNCEGAAGSTSEKSRNPVPAALPVTQQMTEMSISREDKITTPKT-E 118

Query:      116 ETEDVQASVSDTAQQPSEEQS--KPLEKPKQKKNRCFMCRKKVGLTGFECCRCGNVYCGVH 173
             +E V S + QPS QS K E PK KKNRCFMCRKKVGLTGFECCRCGN++CG+H
Sbjct:      119 VSEPVVVTQPSPSVSQPSSSQSEKAPLPKPKKNRCFMCRKKVGLTGFECCRCGNLFCGLH 178

Query:      174 RYSDVLNCSYNYKADAAEKIRKENPVVVGEKIQKI 208
             RYSD NC Y+YKA+AA KIRKENPVVV EKIQ+I
Sbjct:      179 RYSDKHNCYPDYKAEAAAKIRKENPVVVAEKIQR 213

```

Pedant information for DKFZphfbr2_16f21, frame 1

Report for DKFZphfbr2_16f21.1

```

[LENGTH]      208
[MW]           22541.23
[pI]           6.80
[HOMOL]        TREMBL:AF062072_1 gene: "ZNF216"; product: "zinc finger protein 216"; Homo
sapiens zinc finger protein 216 (ZNF216) gene, complete cds. 9e-57
[PIRKW]        zinc 8e-13
[PIRKW]        zinc finger 8e-13

```

[PIRKW] fusion protein 8e-13
 [SUPFAM] unassigned ubiquitin-related proteins 8e-13
 [SUPFAM] ubiquitin homology 8e-13
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] ASN_GLYCOSYLATION 4
 [KW] Irregular
 [KW] LOW_COMPLEXITY 7.21 %

SEQ MAQETNHSQVPMCLSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPPATSVSSLSE
 SEG
 PRD cccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccccc
 SEQ SLPVQCTDGSVPEAQSALDSTSSSMQSPVSNQSLLESVASSQLDSTSVDKAVPETEDV
 SEGXXXXXXXXXXXXXXXXX.....
 PRD ccc
 SEQ QASVSDTAQQPSEEQSKPLEKPKQKKNRCFCMRKKVGLTGFECCRCGNVYCGVHRYSDVLN
 SEG
 PRD ccccccccccccccccccccccccccccccccccecccccccccecccccccccccccccccc
 SEQ CSYNYKADAAEKIRKENPVVVGEKIQKI
 SEG
 PRD ccchhhhhhhhhhhhhcccccccccccccc

Prosite for DKFZphfbr2_16f21.1

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	42->46	ASN_GLYCOSYLATION	PDOC00001
PS00001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS00001	180->184	ASN_GLYCOSYLATION	PDOC00001
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	70->74	CK2_PHOSPHO_SITE	PDOC00006
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	166->172	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16f21.1)

DKFZphfbr2_16g18

group: cell cycle

DKFZphfbr2_16g18.3 encodes a novel 984 amino acid protein with similarity to centromeric proteins of yeasts.

The novel protein shows similarity to *S. pombe* SPAC17A5.07c and the *S. cerevisiae* Smt4p suppressor of MIF2 gene. MIF2 encodes a centromeric protein with homology to the mammalian centromeric protein CENP-C. Mutations in MIF2 stabilise dicentric minichromosomes and confer high instability to chromosomes that bear a cis-acting mutation in element I of the yeast centromeric DNA (CDEI). Therefore the new protein should be involved in centromer organisation, too.

The new protein can find application in modulating/blocking the cell cycle and influencing the behavior of chromosomes, both natural and artificial in eukaryotic cells.

similarity to KIAA0797 and yeast Smt4p

complete cDNA, complete cds, EST hits
the yeast Smt4 protein seems to be involved in centromer function
and microtubule organisation

Sequenced by Qiagen

Locus: unknown

Insert length: 4826 bp

Poly A stretch at pos. 4756, polyadenylation signal at pos. 4736

```

1 GGGTCGAGGT CGACGGTATC GATAAGTTT TTTTTTTTTT TTTTTTTTTT
51 TTTTCCTTTC CCCTCCCCCT CCCTCTCCAA GCCGGAGGGG TCCTGAGGTG
101 ACAGCGCCTG CAACTGAAAT TTCAGCAGCG GGAGAAGATG GACAAGAGAA
151 AGCTCGGGCG ACGGCCATCT TCATCCGAAA TCATCACAGA AGGAAAAAGG
201 AAAAAAGTCAT CTTCTGATTT ATCGGAGATA AGAAAGATGT TAAATGCAAA
251 ACCAGAGGAT GTCCATGTTC AATCACCAC GTCCAAATTC AGAAGCTCAG
301 AACGCTGGAC TCTCCCTTTG CAGTGGGAAA GAAGCCTAAG GAATAAAGTC
351 ATCTCTCTAG ACCATAAAAA TAAAAACAT ATCCGAGGGT GTCCTGTTAC
401 TTCCAGGTCA TCACCAGAAA GGATACCCAG AGTTATATTG ACGAATGTCC
451 TGGGAACGGA GTTAGGAAGA AAATACATAA GGACCCACC TGTAAGTGA
501 GGAAGTTTGA GTGATACAGA CAACTTGCAA TCAGAGCAAC TTTCTTCATC
551 ATCTGATGGC AGCCTAGAAT CTTATCAAAA TCTAAACCTT CACAAGAGCT
601 GTTATTTATC TGAAAGGGGC TCACAACGAA GTAAGACAGT AGATGACAAT
651 TCTGCAAAAGC AGACTGCGCA CAATAAAGAA AAACGAAGAA AGGATGATGG
701 CATTTCTCTT TTAATATCTG ATACTCAGCC TGAAGACCTT AACAGTGGAA
751 GTAGAGGTTG TGATCATCTC GAACAGGAAA GCAGAAACAA GGATGTTAAA
801 TATTCGTGAT CAAAAGTGGG ACTCACTCTG ATTTCCAGGA AGACAAGAG
851 AAGGCTTAGA AATAATTTAC CTGATTCTCA ATATTGTACT TCTTTGGATA
901 AGTCAACAGA ACAGACAAAA AAACAAGAAG ATGACTCAAC AATATCCACT
951 GAGTTTGAAA GGCCAAGTGA AAATATCAT CAGGATCCAA AACTGCCTGA
1001 AGAAATTACA ACTAAACCTA CAAAAGTGA TTTACTAAG CTATCCTCAC
1051 TTAACAGTCA GGAGTTGACT TTGAGTAATG CCACCAAAAG TGCCCTGTC
1101 GGTTCACACA CTGAAACCGT TGAGTACTCT AATTCCATTG ATATTGTGGG
1151 GATTTCTTCC CTGGTTGAGA AGGATGAGAA TGAGTTGAAT ACCATAGAAA
1201 AGCCTATTCT AAGAGGACAT AATGAAGGGA ACCAATCACT GATCTCAGCT
1251 GAACCAATTG TTGTTTCCAG TGATGAAGAA GGACCTGTTG AACATAAAAG
1301 TTCAGAAATT CTAAAGTTAC AATCTAAGCA AGACCGTGAG ACAACTAATG
1351 AAAATGAGAG TACTTCTGAA TCAGCATTGT TAGAACTACC ATTGATTACA
1401 TGTGAATCTG TACAGATGTC ATCTGAATTA TGCCCATATA ATCCTGTCAT
1451 GGAGAACATT TCCAGTATTA TGCCTAGTAA TGAGATGGAT CTACAACCTG
1501 ATTTTATATT TACTTCTGTT TATATTGGTA AAATAAAAGG AGCTTCTAAA
1551 GGTGTGTGTA CAATCACAAA AAAATATATT AAGATCCCAT TTCAAGTGTC
1601 CCTGAATGAG ATTTCATTGC TAGTGGATAC CACACATTTA AAGCGGTTTG
1651 GGTATGGAAG AAGTAAGGAT GATAATCACA GTAAAAGGAG TCATGCTATT
1701 CTTTCTCTCT GGGTCTCTTC AGATTATCTT CAAGAGATTC AGACCCAATT
1751 AGAAGCACTC GTATTAAAGC AGCAATCAAA ATCTAGTGAA TTCATTTTCC
1801 TTGAAGTACA CAATCCTGTT TCACAGAGAG AAGAATTGAA GCTGAAAGAT
1851 ATTATGACCG AAATAAGTAT AATCAGTGGA GAATTAGAGC TTTCTTACCC
1901 GTTGTCTTGG GTTCAGGCAT TTCCTTTGTT TCAGAACCTC TCTTCAAAAG
1951 AAAGTTCTTT TATTCATTAT TACTGTGTTT CAACTGTGTC TTTCCCTGCT
2001 GGTGTTGCTG TTGCTGAAGA AATGAAGCTG AAATCAGTAT CTCAGCCCTC
2051 AAACACAGAT GCGGCCAAGC CTACTTACAC CTTCTGCAG AAGCAAAGTA
2101 GCGGTTGCTA CTCCCTTTCT ATTACATCTA ATCCAGATGA AGAATGGCGG
2151 GAAGTCAGCG ACACTGGACT TGTTCAGAAG TTGATTGTAT ATCCTCCACC
2201 ACCTACTAAG GGGGATTTGG GAGTAACTAA TGAAGATCTG GAGTGTTTGA
2251 AAGAAGGAGA GTTCTTAAAT GATGTAATCA TTGATTTTAA CCTTAAGTAT
2301 CTTATATTGG AGAAGGCATC AGATGAACCT GTTGAACGAA GTCACATTTT

```

```

2351 TAGTAGCTTT TTCTATAAAT GCTTGACAAG AAAGGAAAAT AATTTAACAG
2401 AAGATAATCC AAATCTTTCA ATGGCACAGA GAAGACATAA AAGAGTAAGA
2451 ACATGGAGTC GTCACATAAA CATTTTTAAT AAAGATTACA TCTTTGTACC
2501 TGTAATGAG TCGTCTCACT GGTATCTCGC AGTCATTTGT TTTCCATGGT
2551 TAGAAGAAGC TGTGTATGAA GATTTTCCAC AAAGTGTATC CCAGCAGTCC
2601 CAGGCTCAGC AGTCCCAAAG TGACAACAAA ACAATAGATA ATGATCTACG
2651 TACTACTTCG ACACGTCTTT TGAGTGCAGA GGATTTCCAA AGTACCGAGT
2701 CGAATATGTC AGTACCAAAG AAAATGTGTA AAAGGCCATG TATTCTTATA
2751 CTAGACTCCT TGAAAGCTGC TTCTGTACGA AACACAGTTC AGAATTTACG
2801 AGAGTATTTA GAGGTAGAGT GGGAAAGTTAA ACTAAAACT CATCGTCAAT
2851 TCAGCAAAC AAACATGGTG GATCTATGCC CTAAAGTTCC TAAACAGGAC
2901 AATAGCAGTG ATTTGGGAGT ATATTATTG CAGTATGTGG AAAGCTTCTT
2951 CAAGGATCCT ATTTGTTAACT TTGAACCTCC AATTCATTTG GAGAAGTGGT
3001 TTCTCTGTC TGAATAAAG ACCAAACGGG AAGATATTGC AGAGCTCATC
3051 TTGAACCTTC ATTTACAGCA ACAGAAGGGC AGCAGTAGCT AGTTAATCTG
3101 TACAAACATG ACACAGATGT TCTCTAAGAT TACTGGAAAG CCCCTTACCA
3151 GCATTTGTGT TAGCCAGCTC ACAGAGAAGA AAATAACTTG CAGTAGTTTT
3201 ATAATAAGTC ATTTGAACAT TATTTAAAT ATGTAGGACA CATTATTAGA
3251 ATTTGTTGGA TCTCATAGAT GGAATGGGAA TGGGGGTGAT ATAGATAAAC
3301 TTACTAGATA TAAATTAAAA TTTTATAAAT ATTTCATATT TTTCTGAGTA
3351 AATATGATTG GATTATGCAA CAGCATATGT AATATGGGAA TGTTTTGTAG
3401 ATAATAAAC TTACATGATC TGTACTTCCA CGTACTGGG TGCTGAGGGG
3451 AGTTAAAGCC TCCCTGGTGC CAGCCCCAGT GCTTGTCAA TTTGCTGACA
3501 GGTCACATCA TATTGTAATT CTATCTTTG CAGCTCAAGC ATGCAGTATG
3551 AATACTGTGT ATTTTAAAA AAAATAATTT AGTATCAAGG CTTACAGAAA
3601 TGCCATTTAC GGCATCCCTT CTGTATGTAA CAAAAAGACA TTCATAATGT
3651 TAGGAAGATG ATAAAAATTC GCTCTTTAA AGTGCAGCTT ATTATCTCA
3701 ATTGCTAAAT ACGATTACTC TGCTTTTTTT TTTTCATTTC TTTTGATGTC
3751 ATATGTGAGT ATCTTATAAT TAGTTTCATT TGTTCAGGGT AAAATTTGAA
3801 ACAAAAAATT TTACCTGTGC AAAATAGTTT TTTAAAAATT ATACATGTAG
3851 CTCAACTTGA GGTACTGCTA TATAAATATT CACTCACATT ATCAGGGAAT
3901 TTATGTATAG TTTCTCTAAT ATAGAAGATA AAATTGGTGT CCTCATAACT
3951 TTAACAAAGA AAACCCCTCAG TCCTATTTAT TAATGGGTAG AATTAAATAT
4001 ATAATTTTAT AGCTCAGTTT ACCCAGTATT CATCTGCAAA GCCAGATTGC
4051 TCTCATTGCT TTTATATTTT TAAATTTGTAG CTTTATAGAGA CCTATGATCC
4101 TCATGGAAC TAATTTTTTA TAAATATTC AGGTAACAGT TCTGAATTCA
4151 TGTGATAATG GTGGCATTAT ATATGATTAA ACACTTCAGA ACTTTCTAAT
4201 GTTATCAGGA GTATTTTGAG GGAGATATGA TTATATTGTA TTTTCTCAGA
4251 TAAGAAAAAT GTTTTTTAAC AATATTATTT TAATCTGTTT TAAGCATCTC
4301 TTAGATTTAC ATTATAACTA CATAAAGCAG TGAAGCAAAG GCAAATTAAG
4351 ATAAAGCTAG AAAGTCTGAA CATTTTATTT CAAATCATA CGAATCGGGG
4401 TCAGTTAAGC CTCAGTATTC TTAGCTTTTG TTGATTTTGG CACTATCTTT
4451 ATATTATTAA ATATATTTGT TGTTTGGATA TTTTATATAA AGATGGCTAT
4501 AATTACATAT TTCATTCCCA ATTTGTGTGT GTTGGGGGGT ACTTTTAAAG
4551 GTGACTATTG TTTTGTACAT CTAATTTTGG GAAACCAAGT CTATAAGACA
4601 TCTTGTGATT TCTTAATGTT TTTGTTTGTG TGTTTTCAA AGATATCACT
4651 GTCCTTTATC ATGTTTGAAG GATTGTTTAA AATTCATTTT CCTAAATTA
4701 TGTGCAAGTA ATGTTTGAAG GATATCGGTG TTTTATATTA AACATATTTT
4751 CAATTCAAAA AAAAAAATAA AAAAACTTAT CGATACCGTC GACCTCGATG
4801 ATGATGATGA TGATGATGAT GTCGAC

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 138 bp to 3089 bp; peptide length: 984
 Category: similarity to known protein

```

1 MDKRLKLRFP SSSEIITEGK RKKSSSDLSE IRKMLNAKPE DVHVQSPLSK
51 FRSSERWTLF LQWERSLRNK VISLDHKNKK HIRGCPVTSR SSPERIPRVI
101 LTNVLGTLEL RKYIRTPPVV EGSLSDTDNL QSEQLSSSSD GSLESYQNLN
151 PHKSCYLSER GSQRSKTVDD NSAKQTAHNK EKRRKDDGIS LLISDTQPED
201 LNSGSRGCDH LEQESRNKDV KYSDSKVELT LISRKTNRRL RNNLPDSQYC
251 TSLDKSTEQT KKQEDDSTIS TEFERPSENY HQDPKLPEEI TTKPTKSDFT
301 KLSLNSQEL TLSNATKSAS AGSTTETVEY SNSIDIVGIS SLVEKDENEL
351 NTIEKPILRG HNEGNSQLIS AEPVIVSSDE EGPVEHKSSE ILKLQSKQDR
401 ETTNENESTS ESALLEPLI TCESVQMSSE LCPYNPVMEN ISSIMPSNEM
451 DLQLDFIFTS VYIGKIKGAS KGCVTITKKY IKIPFQVSLN EISLLVDTHH

```

```

501 LKRFGWLKSK DDNHSKRSHA ILFFWSSDY LQEIQTQLEH SVLSQSKSS
551 EFIFLELHNP VSQREELKLG DIMTEISIIS GELELSYPLS WVQAFPLFQN
601 LSSKESFIH YYCVSTCSFP AGVAEEMK LKSVSQPSNT DAAKPTYTFL
651 QKQSSGCYSL SITSNPDEEW REVRHTGLVQ KLIVYPPPTT KGGGLVTNED
701 LECLEEGEFL NDVIIDFYLK YLILEKASDE LVERSHIFSS FFYKCLTRKE
751 NNLTEDNPNL SMAQRRHKRV RTWTRHINIF NKDYIFVPVN ESSHWYLAVI
801 CFPWLEEAVY EDFPQTVSQQ SQAQSQSDN KTIDNDLRTT STLSLSAEDS
851 QSTESNMSVP KKMCKRPCIL ILDSLKAASV RNTVQNLREY LEVEWEVKLK
901 THROFSKTNM VDLCPKVPKQ DNSSDCGVYL LQYVESFFKD PIVNFELPIH
951 LEKWFPRHVI KTKREDIREL ILKLHLQQQK GSSS

```

BLASTP hits

Entry SPAC17A5.7 from database TREMBL:
 "SPAC17A5.07c"; product: "hypothetical protein"; S.pombe
 chromosome I cosmid c17A5. Schizosaccharomyces pombe (fission
 yeast)
 Length = 652
 Score = 275 (96.8 bits), Expect = 1.9e-29, Sum P(3) = 1.9e-29
 Identities = 56/120 (46%), Positives = 78/120 (65%)

Entry S49947 from database PIR:
 SMT4 protein - yeast (Saccharomyces cerevisiae)
 Length = 1034
 Score = 163 (57.4 bits), Expect = 4.6e-16, Sum P(3) = 4.6e-16
 Identities = 46/159 (28%), Positives = 76/159 (47%)

Entry YQG6_CAEEL from database SWISSPROT:
 HYPOTHETICAL 35.7 KD PROTEIN C41C4.6 IN CHROMOSOME II.
 Length = 342
 Score = 162 (57.0 bits), Expect = 6.1e-13, Sum P(3) = 6.1e-13
 Identities = 37/119 (31%), Positives = 62/119 (52%)

Entry AB018340.1 from database TREMBL:
 gene: "KIAA0797"; product: "KIAA0797 protein"; Homo sapiens mRNA for
 KIAA0797 protein, partial cds.
 Score = 540, P = 1.9e-50, identities = 120/243, positives = 155/243

Alert BLASTP hits for DKFZphfbr2_16gl8, frame 3

TREMBL:ATT16L1.11 gene: "T16L1.110"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII
 project), N = 2, Score = 239, P = 2.1e-18

>TREMBL:ATT16L1.11 gene: "T16L1.110"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII project)
 Length = 710

HSPs:

Score = 239 (35.9 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 51/135 (37%), Positives = 78/135 (57%)

Query: 683 IVYPPPTKGGGLGVTNEDLECLEEGEFLNDVIIDFYLYLILEKASDELVERSHIFSSFF 742
 +VYP + V +D+E L+ F+ND IIDFY+KYL + S + R H F+ FF
 Sbjct: 176 LVYPQGEPAVV-VRKQDIELLKPRRFINDTIIDFYIKYL-KNRISPKERGRFHFNCFF 233

Query: 743 YKCLTRKENNLTEDNPNLSMAQRRHKRVRTWTRHINIFNKDYIFVPVNESSHWYLAVICF 802
 + RK NL + P+ + ++RV+ WT+++++F KDYIF+P+N S HW L +IC
 Sbjct: 234 F----RKLANLDKGTPTSCGGREAYQRVQKWTKNVDLFEDKYIFIPINCSFHWLSLVIICH 289

Query: 803 PWLEEAVYEDFPQTV 817
 P + + PQ V
 Sbjct: 290 PGELVPSHVENPQRV 304

Score = 70 (10.5 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 13/28 (46%), Positives = 15/28 (53%)

Query: 948 PIHLEKWFPRHVIKTKREDIRELILKLH 975
 P HL WFP KR +I EL+ LH
 Sbjct: 403 PSHLRNWFPAKEASLKRRIELLYNLH 430

Pedant information for DKFZphfbr2_16gl8, frame 3

Report for DKFZphfbr2_16gl8.3

(LENGTH) 984
 (MW) 112265.80
 (pI) 6.13
 (HOMOL) TREMBL:AB018340_1 gene: "KIAA0797"; product: "KIAA0797 protein"; Homo sapiens
 mRNA for KIAA0797 protein, partial cds. 8e-53
 (FUNCAT) 03.22 cell cycle control and mitosis [S. cerevisiae, YIL031w] 9e-17
 (FUNCAT) 99 unclassified proteins [S. cerevisiae, YPL020c] 4e-06
 (BLOCKS) BL00494C Bacterial luciferase subunits proteins
 (PROSITE) AMIDATION 3
 (PROSITE) MYRISTYL 9
 (PROSITE) CAMP_PHOSPHO_SITE 2
 (PROSITE) CK2_PHOSPHO_SITE 30
 (PROSITE) TYR_PHOSPHO_SITE 1
 (PROSITE) PKC_PHOSPHO_SITE 19
 (PROSITE) ASN_GLYCOSYLATION 12
 (KW) Alpha_Beta
 (KW) LOW_COMPLEXITY 4.47 %

SEQ MDKRLGRRPSSSEIITEGKRKKSSSDLSEIRKMLNAKPEDVHVQSPLSKFRSSERWTLP
 SEG
 PRD cch

SEQ LQWERSLRNKVISLDHKNKKHIRGCPVTSRSSPERIPRVILTNLVGTGLGRKYIIRTPPVT
 SEG
 PRD hhhhhhhhhheeeccccceeecc

SEQ EGSLSDTNDLQSEQLSSSSDGSLESYQNLNPHKSCYLSESGRSQRSKTVDDNSAKQTAHNK
 SEGxxxxxxxxxxxxxxxxxxxxx.....
 PRD cchhhhhhh

SEQ EKRRKDDGISLLISDTQPEDLNSGSRGCDHLEQESRNKDKVYSDSKVELTLISRKTKRRL
 SEG
 PRD hhhhccchhhhhhh

SEQ RNNLPDSQYCTSLDKSTEQTQKQEDDSTISTEFERPSSENYHQDPKLPKEITTKPTKSDFT
 SEG
 PRD hccccccccccccccccccccchhhhhcccccccccccccccccccccccccccccccccccc

SEQ KLSSLNSQELTSLNATKSASAGSTTETVEYSNSIDIVGISSLVEKDENEELNTIEKPILRG
 SEG
 PRD cccccccccceehhhhhhhcc

SEQ HNEGQSLISAEPDIVSSDEEGPVEHKSSEILKLSQKQDRETTNENESTSESALLEPLI
 SEGxxxxxxxxxxxxxxxxxxxxx.....
 PRD cccccceeeccccccccccccccccchhhhhhhhhhhhhccccccccchhhhhccccce

SEQ TCESVQMSELCPYNPVMEISSIMPSNEMDLQLDFIFTSVYIGIKIGASKGCVTITKKY
 SEG
 PRD eccccccccccccccccccccccccccccchhhhhhhheeeeeeeeecccccccccccccccc

SEQ IKIPFQVSLNEISLLVDTTHLKRFLGWSKDDNHSKRSHAILFFWVSSDYLQEIQTQLEH
 SEG
 PRD eeeccchhhhhhhhhhh

SEQ SVLSQSKSSEFIFLELHNPVSQREELKLDIMTEISIIISGELELSYPLSWVQAFPLFQN
 SEG
 PRD hhhhccccccccccccccccchhhhhhhhhhhheeecccccccccccccccccccccccc

SEQ LSSKESSFIHYCVSTCSFPAGVAVAEEMKLKSVSQPSNTDAKPTYTFLQKQSSGCYSL
 SEG
 PRD cccccccccceeeccccccccchhhhhhhhhhhcccccccccccccccccccccccccccc

SEQ SITSNPDEEWREVRHTGLVQKLIYVPPPTKGGLVGTNEDLECLEEGFLNDVIIDFYLK
 SEG
 PRD eccccccccceeeccccccccccccccccccccccccchhhhhhhhhccccchhhhhhhhhhh

SEQ YLILEKASDELVERSHIFSSFFYKCLTRKENNLTEDNPNLSMAQRRHKRVRTWTRHINIF
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhc

SEQ NKDYIFVPVNESSHWYLAVICFPWLEEAVYEDFPQTVSQQSQQQSQSDNKTIDNDLRTT
 SEGxxxxxxxxxxxxx.....
 PRD cceeeccccccccccccccccchhhhhhhccccchhhhhhhhhhhcccccccccccccccc

SEQ STLSSLASDSQSTESNMSVPKMKCRPCILILDSLKAASVRNTVQNLREYLEVEWEVKLK
 SEG
 PRD cceeeccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhh

SEQ THRQFSKTNMVDLCPKVPKQDNSSDCGVYLLQYVESFFKDPIVNFELPIHLEKWFPRHVI

```

SEG      .....
PRD      hhhhhccccccccccccccccceeeehhhhhhhcccceccccccccccchhh

SEG      KTKREDIRELILKLHLQOKGSSS
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhccccc

```

Prosites for DKFZphfbr2_16g18.3

PS000001	314->318	ASN_GLYCOSYLATION	PDOC000001
PS000001	365->369	ASN_GLYCOSYLATION	PDOC000001
PS000001	406->410	ASN_GLYCOSYLATION	PDOC000001
PS000001	440->444	ASN_GLYCOSYLATION	PDOC000001
PS000001	513->517	ASN_GLYCOSYLATION	PDOC000001
PS000001	600->604	ASN_GLYCOSYLATION	PDOC000001
PS000001	752->756	ASN_GLYCOSYLATION	PDOC000001
PS000001	759->763	ASN_GLYCOSYLATION	PDOC000001
PS000001	790->794	ASN_GLYCOSYLATION	PDOC000001
PS000001	830->834	ASN_GLYCOSYLATION	PDOC000001
PS000001	856->860	ASN_GLYCOSYLATION	PDOC000001
PS000001	922->926	ASN_GLYCOSYLATION	PDOC000001
PS000004	8->12	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	21->25	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	54->57	PKC_PHOSPHO_SITE	PDOC000005
PS000005	66->69	PKC_PHOSPHO_SITE	PDOC000005
PS000005	88->91	PKC_PHOSPHO_SITE	PDOC000005
PS000005	158->161	PKC_PHOSPHO_SITE	PDOC000005
PS000005	162->165	PKC_PHOSPHO_SITE	PDOC000005
PS000005	172->175	PKC_PHOSPHO_SITE	PDOC000005
PS000005	233->236	PKC_PHOSPHO_SITE	PDOC000005
PS000005	236->239	PKC_PHOSPHO_SITE	PDOC000005
PS000005	260->263	PKC_PHOSPHO_SITE	PDOC000005
PS000005	291->294	PKC_PHOSPHO_SITE	PDOC000005
PS000005	477->480	PKC_PHOSPHO_SITE	PDOC000005
PS000005	515->518	PKC_PHOSPHO_SITE	PDOC000005
PS000005	562->565	PKC_PHOSPHO_SITE	PDOC000005
PS000005	602->605	PKC_PHOSPHO_SITE	PDOC000005
PS000005	747->750	PKC_PHOSPHO_SITE	PDOC000005
PS000005	874->877	PKC_PHOSPHO_SITE	PDOC000005
PS000005	879->882	PKC_PHOSPHO_SITE	PDOC000005
PS000005	901->904	PKC_PHOSPHO_SITE	PDOC000005
PS000005	962->965	PKC_PHOSPHO_SITE	PDOC000005
PS000006	11->15	CK2_PHOSPHO_SITE	PDOC000006
PS000006	24->28	CK2_PHOSPHO_SITE	PDOC000006
PS000006	91->95	CK2_PHOSPHO_SITE	PDOC000006
PS000006	123->127	CK2_PHOSPHO_SITE	PDOC000006
PS000006	125->129	CK2_PHOSPHO_SITE	PDOC000006
PS000006	137->141	CK2_PHOSPHO_SITE	PDOC000006
PS000006	167->171	CK2_PHOSPHO_SITE	PDOC000006
PS000006	196->200	CK2_PHOSPHO_SITE	PDOC000006
PS000006	225->229	CK2_PHOSPHO_SITE	PDOC000006
PS000006	251->255	CK2_PHOSPHO_SITE	PDOC000006
PS000006	271->275	CK2_PHOSPHO_SITE	PDOC000006
PS000006	295->299	CK2_PHOSPHO_SITE	PDOC000006
PS000006	323->327	CK2_PHOSPHO_SITE	PDOC000006
PS000006	341->345	CK2_PHOSPHO_SITE	PDOC000006
PS000006	377->381	CK2_PHOSPHO_SITE	PDOC000006
PS000006	396->400	CK2_PHOSPHO_SITE	PDOC000006
PS000006	402->406	CK2_PHOSPHO_SITE	PDOC000006
PS000006	408->412	CK2_PHOSPHO_SITE	PDOC000006
PS000006	488->492	CK2_PHOSPHO_SITE	PDOC000006
PS000006	509->513	CK2_PHOSPHO_SITE	PDOC000006
PS000006	536->540	CK2_PHOSPHO_SITE	PDOC000006
PS000006	562->566	CK2_PHOSPHO_SITE	PDOC000006
PS000006	602->606	CK2_PHOSPHO_SITE	PDOC000006
PS000006	638->642	CK2_PHOSPHO_SITE	PDOC000006
PS000006	664->668	CK2_PHOSPHO_SITE	PDOC000006
PS000006	697->701	CK2_PHOSPHO_SITE	PDOC000006
PS000006	747->751	CK2_PHOSPHO_SITE	PDOC000006
PS000006	826->830	CK2_PHOSPHO_SITE	PDOC000006
PS000006	846->850	CK2_PHOSPHO_SITE	PDOC000006
PS000006	962->966	CK2_PHOSPHO_SITE	PDOC000006
PS000007	216->223	TYR_PHOSPHO_SITE	PDOC000007
PS000008	84->90	MYRISTYL	PDOC000008
PS000008	106->112	MYRISTYL	PDOC000008
PS000008	141->147	MYRISTYL	PDOC000008
PS000008	161->167	MYRISTYL	PDOC000008
PS000008	204->210	MYRISTYL	PDOC000008
PS000008	468->474	MYRISTYL	PDOC000008

PS00008	505->511	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00008	693->699	MYRISTYL	PDOC00008
PS00009	6->10	AMIDATION	PDOC00009
PS00009	18->22	AMIDATION	PDOC00009
PS00009	109->113	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_16g18.3)

DKFZphfbr2_16i12

group: transmembrane protein

DKFZphfbr2_16i12 encodes a novel 185 amino acid protein, with strong similarity to PUT2 protein of Fugu rubripes.

The novel protein contains 1 transmembrane region.

PUT 2 is a Fugu rubripes protein similar to the neural cell adhesion molecule L1 (L1-CAM) a mitosis-specific chromosome segregation protein (SMC1) and the calcium channel alpha-1 subunit homolog (CCA1).

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

strong similarity to Fugu rubripes PUT2

complete cdna, complete cds, EST hits,
TRANSMEMBRANE 1

Sequenced by LMU

Locus: /map="873.3/875.1 cR from top of Chr1 linkage group"

Insert length: 1552 bp

Poly A stretch at pos. 1528, polyadenylation signal at pos. 1506

```
1 GGGGGGGGAC AACTGGGTCT TTTGCGGCTG CAGCGGGCTT GTAGGCGTCC
51 GGCTTTGCTG GCCCAGCAAG CCTGATAAGC ATGAAGCTCT TATCTTTGGT
101 GGCTGTGGTC GGGTGTTTGC TGGTGCCCCC AGCTGAAGCC AACAAAGAGTT
151 CTGAAGATAT CCGGTGCAAA TGCATCTGTC CACCTTATAG AAACATCAGT
201 GGGCACATTT ACAACCAGAA TGTATCCAG AAGGACTGTT GTAGCAACTG
251 CCTGCACGTG GTGGAGCCCA TGCCAGTGCC TGGCCATGAC GTGGAGGCCT
301 ACTGCCCTGCT GTGCGAGTGC AGGTACGAGG AGCGCAGCAC CACCACCATC
351 AAGGTATCA TTGTCATCTA CCTGTCCGTG GTGGGTGCCC TGTGTCTCTA
401 CATGGCCTTC CTGATGCTGG TGGACCTCT GATCCGAAAG CCGGATGCAT
451 ACATGAGCA ACTGCACAAT GAGGAGGAGA ATGAGGATGC TCGCTCTATG
501 GCAGCAGCTG CTGCATCCCT CGGGGGACCC CGAGCAAACA CAGTCTTGGA
551 GCGTGTGGAA GGTGCCCAGC AGCGGTGAA GCTGCAGGTG CAGGAGCAGC
601 GGAAGACAGT CTTGATCGG CACAAGATGC TCAGCTAGAT GGGCTGGTGT
651 GGTGTTGGTCA AGGCCCCAAC ACCATGGCTG CCAGCTTCCA GGCTGGACAA
701 TGGGTCTTTG GGGTTGAAGG GAGGGGGAAG GCAGGCCAGA AGGGAATGGA
751 CCGTGGCAT TTTTCTCTCT TCTCCCTAAC TTTAGAAATG TTGTACTTGG
801 CTATTTTGTAT TAGGGAAGAG GGATGTGGTC TCTGATCTCT GTTGTCTTCT
851 TGGGTCTTTG GGGTTGAAGG GAGGGGGAAG GCAGGCCAGA AGGGAATGGA
901 GACATTCCAG GCGGCCTCAG GAGTGGATGC GATCTGTCTC TCCTGGCTCC
951 ACTCTTGCCG CCTTCCAGCT CTGAGTCTTG GGAATGTTGT TACCCTTGGA
1001 AGATAAAGCT GGGTCTTCAG GAACTCAGTG TTTGGGAGGA AAGCATGGCC
1051 CAGCATTCAG CATGTGTTC TTTCTGCAGT GGTCTTATC ACCACCTCCC
1101 TCCCAGCCCC AGCGCCTCAG CCCCAGCCCC AGCTCCAGCC CTGAGGACAG
1151 CTCTGATGGG AGAGCTGGGC CCCCTGAGCC CACTGGGTCT TCAGGGTGCA
1201 CTGGAAGCTG GTGTTGCTG TCCCCTGTGC ACTTCTCGCA CTGGGGCATG
1251 GAGTGCCCAT GCATACTCTG CTGCCGTGCC CCTCACCTGC ACTTGAGGGG
1301 TCTGGGCAGT CCCTCCTCTC CCCAGTGTCC ACAGTCACTG AGCCAGACGG
1351 TCGGTTGGAA CATGAGACTC GAGGCTGAGC GTGGATCTGA ACACCACAGC
1401 CCCTGTACTT GGGTTGCCTC TTGTCCCTGA ACTTCGTTGT ACCAGTGCAT
1451 GGAGAGAAAA TTTTGTCTCT TTGTCTTAGA GTTGTGTGTA AATCAAGGAA
1501 GCCATCATTA AATTGTTTA TTTCTCTCAA AAAAAAAAAA AAAAAAATA
1551 TC
```

BLAST Results

Entry HS808349 from database EMBL:
human STS WI-11986.
Score = 1716, P = 5.7e-73, identities = 364/378

Entry HS487355 from database EMBL:
human STS WI-13088.
Score = 1358, P = 1.3e-56, identities = 274/277

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 81 bp to 635 bp; peptide length: 185
Category: similarity to unknown protein

```

1  MKLLSLVAVV GCLLVPPAEA NKSSDIRCK CICPPYRNIS GHIYNQNVSQ
51 KDCCSNCLHV VEPMPVPGHD VEAYCLCEC RYEERSTTTI KVIIVIVLSV
101 VGALLLYMAF LMLVDPLIRK PDAYTEQLHN EEENEDARSM AAAAASLGGP
151 RANTVLERVE GAQQRWKLQV QEQRKTVFDR HKMLS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16i12, frame 3

TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu
rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene,
complete cds; putative protein 1 (PUT1) gene, partial cds;
mitosis-specific chromosome segregation protein SMC1 homolog (SMC1)
gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1)
and putative protein 2 (PUT2) genes, partial cds, complete sequence., N
= 1, Score = 655, P = 2.8e-64

TREMBL:CER12C12_5 gene: "R12C12.6"; Caenorhabditis elegans cosmid
R12C12., N = 1, Score = 225, P = 1e-18

>TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu
rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete
cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific
chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and
calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2
(PUT2) genes, partial cds, complete sequence.
Length = 187

HSPs:

Score = 655 (98.3 bits), Expect = 2.8e-64, P = 2.8e-64
Identities = 124/163 (76%), Positives = 140/163 (85%)

```

Query: 22 KSSDIRCKCICPPYRNISGHIYNQNVSQKDCCSNCLHVVEPMPVPGHDVEAYCLCECR 81
      KS +D+RCKCICPPYRNISGHIYN+N +QKDC NCLHVV+PMPVPG+DVEAYCLCEC+
Sbjct: 31 KSFDVRCCKCICPPYRNISGHIYNRNFTQKDC--NCLHVVDPMVPGNDVEAYCLCECK 88

Query: 82 YEERSTTTIKVIIVIVLSVVGALLLYMAFLMLVDPLIRKPDAYTEQLHNEEENEDARSM 141
      YEERST TI+V I+I+LSVVGALLLYM FL+LVDPLIRKPD + LHNEE++ED +
Sbjct: 89 YEERSTNIRVTIIIFLSVVGALLLYMLFLLLDPLIRKPDPLAQLHNEESEDIIQPQM 148

Query: 142 AAAASLGGP-RANTVLERVEGAQQRWKLQVQEQRKTVFDRHKML 184
      + G P R NTVLERVEGAQQRWK QVQEQRKTVFDRHKML
Sbjct: 149 S-----GDPARGNTVLERVEGAQQRWKKQVQEQRKTVFDRHKML 187

```

Pedant information for DKFZphfbr2_16i12, frame 3

Report for DKFZphfbr2_16i12.3

```

[LENGTH] 185
[MW] 20764.29
[pI] 6.21
[HOMOL] TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu rubripes
neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1
(PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1)
gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2
(PUT2) genes, partial cds, complete sequence. 3e-68
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 2
[PROSITE] ASN_GLYCOSYLATION 3
[KW] SIGNAL_PEPTIDE 21

```



```

[KW]          TRANSMEMBRANE 1
[KW]          LOW_COMPLEXITY 2.70 %

SEQ      MKLLSLVAVVGCLLVPPAEANKSSEDIRCKCICPPYRNISGHIYNQNVSQKDCCSNCLHV
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      VEPMPVPGHDVEAYCLLCECRYEERSTTTIKVIIIVYLSVVGALLLYMAFLMLVDPLIRK
SEG      .....
PRD      eccccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      PDAYTEQLHNEEENEDARSMASLGGPRANTVLERVEGAQQRWKLQVQEQRKTVFDR
SEG      .....xxxxx.....
PRD      ccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      HKMLS
SEG      .....
PRD      hhccc
MEM      .....

```

Prosite for DKFZphfbr2_16i12.3

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	38->42	ASN_GLYCOSYLATION	PDOC00001
PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	148->154	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16i12.3)

DKFZphfbr2_16k22

group: brain derived

DKFZphfbr2_16k22 encodes a novel 108 amino acid protein with very weak similarity to thioredoxin of *Bacillus subtilis*.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to thioredoxin

complete cDNA, complete cds, genomic DNA?
no EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2088 bp
Poly A stretch at pos. 2065, no polyadenylation signal found

```
1 AAAAGGAAGA AGGAAATAAG GATATTTCAG GGGTTACCAA AGTCGAGGAA
51 AACTATTTTA AGAAGAAATC TGAATTATTT GTGCACATAG GTTGTAAATA
101 TAGCATCTTG CATTAAATGG TGTTTTCTAG CTTACAAAGT GGATTCATAT
151 ACACATTTGT AACTGACTCT CTACAAACTT GCAAGGTTAG CAAGACAAAT
201 GGTATTTTAA GATAACAAAC TGAGACTCAA AAAAGGCAAG TAACCTGTTT
251 TACTTCCCAA AGCCAGAAAG TGGCAAAATA GAAAATGGAT CCTGAATCTC
301 CAACACCATG CAAACTAAGA GAGGGAATCC TCTGTAGAGG GAATGGAAGT
351 AAAAAGGCAC AAGTGGTGAT GTCACCTTCT GAACAGAGAT GGAACCTTTC
401 TTCCTCTGAG AAAAAGAGA AAAGATAGTT TTAAGTGGCA AAAGAACATG
451 AAGCAATGTG AGGTGAAGAA ACAGAAAAGA CTATGGATGG AATTCCTAGA
501 TGTGAGATAC ACAAGATTCC ATTTCAAAGA GAAATATCTA TAGATAGGCA
551 TAAAGTTACA CACCTGAACCT ACCAACTCTG AACCAGTAAC TCAAGAGATA
601 TTTTGTGTGT CCCACAAGCC ATATGGCTCT GGGGACAAAT TATCTGAAAG
651 TGCCCAATAA GAAAAATATT TGAGGAAGGG GAGTTGGTGA GTGAATGAAT
701 TAAAGGACAT CAGAAAGATA CATTGACTGT TCTCCTTCCC AGGAAACAAA
751 GTGGCTAAGT CAAAACAACG GGCAGCTGTG GGATAGCAAA GAAAAAAGAA
801 CTTCAGGACC CAGGTTCTAG TGAAAGCTAC TATGGAAGTT AGCCACTCAA
851 CTTTGAAGCC AGAGGCTTCT TTCTCTCCTC CCTTCTTATC TTTTCTAGTT
901 TATAGCAAAAT TTATATTGAG CCACTTATTC TTTCTGAATG CTAGTTCCCC
951 TTTAGCATTT CTTTTTCTTC ATTCCCTTTG GACTGGCCCA ATGCTTTGGC
1001 CCCTTATCAA AGCATTTTCT AAGAAACAGT CTGACAGCTC TAATTTGCAT
1051 CTGGTTATGC AAGATGTGGT TAAGAACATG GACTCTGGAG GTAAATACAC
1101 CTTGATTCCA ATTCATTCTC TCATTATTTC ATTCAGCAAA TATTTAGTGA
1151 ACATCTAACA TGTGCTAGGC ACTGTTCTAG TTGCTGAGGA TACAGCTTCA
1201 AACAAATATA GGTCTCTGCA AGGATGCCTT CTCTTACCAC TCCTATTTCAG
1251 CGTAGTATTG GAAGTCCTGG CCAGGGCAAT CAGGCAAGAA AAAGAAATCA
1301 AGGTCATCCA AATAGGAAGA GAGGAAGTCA AACTATCCCT GTTTACAGAC
1351 AACATGATCC TACATCTAGA AAAAAACCCA TTGTCTTAGC CCAAAAGCTT
1401 CTTAGGCTGA TAAACAACCT CAGCAAAAGT TTAGGATACA AAATCCATGT
1451 GCAAAAAACA CTAGCATTTT TATACACCAA CAACAGTCAA GCCGAGATCC
1501 AAATCAGGAA CAAACTCCTA TTCACAATTG CCACAAAAAC AATAGAACAG
1551 GAAACAGCTC AACTAGGAAG GTGAAAGATC TCTACAAGGA GAACTACAAA
1601 CCACTGCTCA CAGAAATCAG AGATGACACA TATAAATGGA AAAACATTCC
1651 ATGATCATGG ATAGGAAGAA TGAATATTAC TGAATGGCT ATACTGTCCA
1701 AAGCAATTTA TAGATTCAAT GCTATTCCTA GTAAACTACC ATTGAGATTT
1751 TTTACAGAAC TAGAAAAAAA AAAAATCTAT TTAAGGCTGG GCGCAGTGGC
1801 TCTCACCTGT AATCCCAGCA CTTTGGGAGG CCGAGATGGG TGGATCACGA
1851 GGTGAGGAGA TGGAAAAACAT CCTGGCTAAC ATGGTGAAAC CCCGTCTCTA
1901 CTAAAAATAC AAAAAATTAG CCAGGCGTGG TGGTGGGCGC CTGTAATCCC
1951 AGCTGCTCGG GAGGCTGAGG CAGGATAATG GTGTGAACCC GGGAGGCAGA
2001 GCTTGCAAGT AGCTGAGATT GCACCACTGC ACTCCAGCCT GAGGGACAGA
2051 GTGAGACTCC ATCTCAAAAA AAAAAAAGAA AAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 832 bp to 1155 bp; peptide length: 108
Category: putative protein

1 MEVSHSTLEP EASFPPFPLS FLVYSKFILS HLFFLNASSP LAFLFLHSLW
51 TGPMLWPLIK AFSKKQSDSS NLHLVMQDVV KNMDSGGKYT LIPIHSLIYS
101 FSKYLVNI

BLASTP hits

Entry B37192 from database PIR:
thioredoxin - Bacillus subtilis Score = 71 (25.0 bits), Expect = 0.040,
P = 0.039
Identities = 16/49 (32%), Positives = 30/49 (61%)

Alert BLASTP hits for DKFZphfbr2_16k22, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_16k22, frame 1

Report for DKFZphfbr2_16k22.1

[LENGTH] 108
[MW] 12281.47
[pI] 8.06
[PROSITE] MYRISTYL 1
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta

SEQ MEVSHSTLEPEASFPPFPLSFLVYSKFILSHLFFLNASSPLAFLFLHSLWTGPMLWPLIK
PRD cccccccccccccccccchhhhhhhhhhhhhhhccccchhhhhhhhhccccchhhhh

SEQ AFSKKQSDSSNLHLVMQDVVKNMDSGGKYTLIPIHSLIYSFSKYLVINI
PRD hhhccccccccceehhhhhccccccccceeeecceeeeccecccccc

Prosite for DKFZphfbr2_16k22.1

PS00001	36->40	ASN_GLYCOSYLATION	PDOC00001
PS00004	64->68	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00006	6->10	CK2_PHOSPHO_SITE	PDOC00006
PS00008	86->92	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16k22.1)

DKFZphfbr2_16112

group: transmembrane protein

DKFZphfbr2_16112 encodes a novel 267 amino acid protein with similarity to gallus gallus putative transmembrane protein E3-16

The novel protein contains one putative transmembrane domain. In chicken, E3-16 is expressed specifically in the inner ear.

No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neurons involved in perception of hearing.

similarity to gallus putative transmembrane protein E3-16

complete cDNA, complete cds, EST hits
potential start at Bp 73 matches kozak consensus PyCCataG
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 2042 bp

Poly A stretch at pos. 2024, polyadenylation signal at pos. 2003

```
1 GGGGGCGGGC GAGGCAGAGA CCGAGGCTGC ACCGGCAGAG GCTGCGGGGC
51 GGACGCGCGG GCCGGCGCAG CCATGGTGAA GATTAGCTTC CAGCCCCGCC
101 TGGCTGGCAT CAAGGGCGAC AAGGCTGACA AGGCGTCGGC GTCGGCCCTT
151 GCGCCGGCCT CGGCCACCGA GATCCTGCTG ACGCCGGCTA GGGAGGAGCA
201 GCCCCACAA CATCGATCCA AGAGGGGGGG CTCAGTGGGC GGCCTGTGCT
251 ACCTGTGCAT GGCATGGTC GTGCTGCTCA TGGCCCTCGT GTTCGCCTCT
301 GTCTACATCT ACAGATACTT CTTCTTGCG CAGCTGGCCC GAGATAACTT
351 CTTCCGCTGT GGTGTGCTGT ATGAGGACTC CCTGTCTCC CAGGTCCGGA
401 CTCAGATGGA GCTGGAAGAG GATGTGAAA TCTACCTCGA CGAGAACTAC
451 GAGCGCATCA ACGTGCCTGT GCCCCAGTTT GCGGGCGGTG ACCCTGCAGA
501 CATCATCCAT GACTTCCAGC GGGGTCTGAC TCGTACCAT GATATCTCCC
551 TGGACAAGTG CTATGTATC GAACTCAACA CCACCATGTG GCTGCCCCCT
601 CGCAACTTCT GGGAGCTCCT CATGAACGTG AAGAGGGGGA CCTACCTGCC
651 GCAGACGTAC ATCATCCAGG AGGAGATGGT GGTACGGGAG CATGTCAAGT
701 ACAAGGAGGC CCTGGGGTCC TTCACTTACC ACCTGTGCAA CGGGAAGAC
751 ACCTACCGGC TCCGGCGCCG GGCAACGCGG AGGCGGATCA ACAAGCGTGG
801 GGCCAAGAAC TGCAATGCCA TCCGCCACTT CGAGAACACC TTCGTGGTGG
851 AGACGCTCAT CTGCGGGGTG GTGTGAGGCC CTCCTCCCCC AGAACCCCTT
901 GCCGTGTTCC TCTTTTCTT TTTCCGGCTG CTCTCTGGCC CTCCTCCTTC
951 CCCCTGCTTA GCTTGTAATT TGGACGCGTT TCTATAGAGG TGACATGTCT
1001 CTCCATTCTT CTCCAACCTT GCCCACCTCC CTGTACCAGA GCTGTGATCT
1051 CTCGGTGGGG GGCCCATCTC TGCTGACCTG GGTGTGGCGG AGGGAGAGGC
1101 GATGCTGCAA AGTGTCTTCT GTGTCCCACT GTCTTGAAGC TGGGCCTGCC
1151 AAAGCCTGGG CCCACAGCTG CACCGGCAGC CCAAGGGGAA GGACCGGTTG
1201 GGGGAGCCGG GCATGTGAGG CCTGGGCAA GGGGATGGGG CTGTGGGGGG
1251 GGGGCGGCAT GGGCTTCAGA AGTATCTGCA CAATTAGAAA AGTCCTCAGA
1301 AGCTTTTCTT TGGAGGGTAC ACTTTCTTCA CTGTCCCTAT TCCTAGACCT
1351 GGGGCTTGAG CTGAGGATGG GACGATGTGC CCAGGGAGGG ACCCACCAGA
1401 GCACAAGAGA AGGTGGCTAC CTGGGGGTGT CCCAGGGAAT CTGTCAAGTC
1451 CTTACGCCCC CCAGCAGGAG CTTGGAGTTT GGGGAGTGGG GATGAGTCCG
1501 TCAAGCAGAA CTGTTCTCTG AGTGGAAACA AAGAAGCAAG GAGCTAGGAC
1551 CCCCAGTCTT GCCCCCCAGG AGCACAAGCA GGTCCCTCT AGTCAAGGCA
1601 GTGGGATGGG CGGCTGAGGA ACGGGGCGAG CAAGGTCACT GCTCAGTCAC
1651 GTCCACGGGG GACGAGCCGT GGGTTCTGCT GAGTAGGTGG AGCTCATTGC
1701 TTTCTCCAAG CTTGGAAGTG TTTTGAAAGA TAACACAGAG GGAAAGGGAG
1751 AGCCACCTGG TACTTGTCCA CCCTGCCTCC TCTGTCTGTA AATTCCATCC
1801 CCCTCAGCTT AGGGGAATGC ACCTTTTCTT CTTTCTTCT CACTTTTGCA
1851 TGTTTTTACT GATCATTGCA TATGCTAACC GTTCTCAGCC CTGAGCCTTG
1901 GAGAGGAGGG CTGTAACGCC TTCAGTCAGT CTCTGGGGAT GAAACTCTTA
1951 AATGCTTTGT ATATTTTCTC AATTAGATCT CTTTCAGAA GTGTCTATAG
2001 AACAATAAAA ATCTTTTACT TCTGAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation using cDNA library subtraction. Molecular cloning and characterization of a gene belonging to a novel multigene family of integral membrane proteins.

Peptide information for frame 1

ORF from 73 bp to 873 bp; peptide length: 267
Category: similarity to known protein

```

1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
51 RGGSVGGVCY LSMGMVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VVPVQFGGDD PADIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRNFW ELLMNVKRGD YLPQTYIIQE
201 EMVTEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRIN KRGAKNCNAI
251 RHFENTFVVE TLICGVV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16112, frame 1

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16)., N = 1, Score = 573, P = 1.4e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N = 1, Score = 559, P = 4.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1, Score = 452, P = 9.1e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).

Length = 262

HSPs:

Score = 573 (86.0 bits), Expect = 1.4e-55, P = 1.4e-55
Identities = 118/264 (44%), Positives = 175/264 (66%)

```

Query:      1 MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGGSVGGVCY 60
             MVK+SF A+A  + A+K  ++      ++L+ P  + + P+      G    C+
Sbjct:      1 MVKVSFNSALA--HKEAANKEEENS-----QVLILPP-DAKEPEDVVVPAGHKRAWCW 50

Query:      61 -LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSL- ----SQVRTQM- 112
             + G+ +L G++  Y+Y+YF Q      + CG+ Y ED LS      +Q+++
Sbjct:      51 CMCFLAGFMLAGVILGGAYLYKYFAFQQ---GGVYFCGIKYIEDGLSLPESGAQLKSARY 107

Query:     113 -ELEEDVKIYLDENYERINVPVPQFGGDDPADIHDFQRLTAYHDISL DKCYVIELNTT 171
             +E++++I +E+ E I+VPVP+F DPADI+HDF R LTAY D+SLDKCYVI LNT+
Sbjct:     108 HTIEQNIQILEEEDVEFISVPVPEFADSDPADIVHDFHRLTAYLDLSL DKCYVIPLNTS 167

Query:     172 IVLPPRNFWELLMNVKRGTYLPQTYIIQEEMVTEHVSDKEALGSFIYHLCNGKDTYRLR 231
             +V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+
Sbjct:     168 VVMPPKNFLELLINIKAGTYLPQSYLIHEQMIVTDRIENV DQLGFFIYRLCRGKETYLQ 227

Query:     232 RRATRRRRINKRGAKNCNAIRHFENTFVVETLIC 264
             R+   + I KR A NC IRHFEN F +ETLIC
Sbjct:     228 RKEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260

```

Pedant information for DKFZphfbr2_16112, frame 1

Report for DKFZphfbr2_16112.1

[LENGTH] 267
[MW] 30223.94

[pI] 8.16
 [HOMOL] SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
 1e-49
 {PROSITE} PRENYLATION 1
 {PROSITE} MYRISTYL 5
 {PROSITE} CAMP_PHOSPHO_SITE 2
 {PROSITE} CK2_PHOSPHO_SITE 3
 {PROSITE} TYR_PHOSPHO_SITE 1
 {PROSITE} PKC_PHOSPHO_SITE 4
 {PROSITE} ASN_GLYCOSYLATION 1
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 15.36 %

SEQ MVKISFQPAVAGIKGDKADKASAPAPASATEILLTPAREEQPPQHRSKRGGVGGVCY
 SEGXX
 PRD cccccccccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccchh
 MEMMM

SEQ LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLYEDSLSSQVRTQMELEEDVKI
 SEGXX
 PRD hhhccccccccccccchhhhhhhhhhhhhhh
 MEM MMM

SEQ YLDENYERINVPVPQFGGDPADIHDFQRGLTAYHDISLDKCYVIELNTTIVLPFRNFW
 SEGXX
 PRD hhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchh
 MEMXX

SEQ ELLMNVKRGTYLPQTYIIQEEMVTEHVSDEALGSFIYHLCNGKDTYRLRRRATRRRIN
 SEGXX
 PRD hhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhh
 MEMXX

SEQ KRGAKNCNAIRHFENTFVVETLICGVV
 SEG xx.....XX
 PRD hhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccc
 MEMXX

Prosites for DKFZphfbr2_16112.1

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	53->59	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PRENYLATION	PDOC00266

(No Pfam data available for DKFZphfbr2_16112.1)

DKFZphfbr2_22f21

group: brain derived

DKFZphfbr2_22f21 encodes a novel 567 amino acid protein with weak similarity to C. elegans cosmid C18C4.5

No informative BLAST results; no predictive prosite, pfam or SCOP motifs

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to C.elegans C18C4.5

EST HSAA6531/HSAA5273/ defines splice variant, or unspliced cDNA additional ~180 Bp at position 250

Sequenced by AGOWA

Locus: /map="311.4 cR from top of Chr14 linkage group"

Insert length: 1910 bp

Poly A stretch at pos. 1887, polyadenylation signal at pos. 1867

```
1 TGGGCCCTTA GCAACGGCCT GGCACGGTT TCCTTGCTGC TGCAGCCCCC
51 GTGGGCTCCT CTTTCCAGT CCTCCACTGC CGGGGCTGGG CCCGGCCGCG
101 GGAAGGACCG AAGGGGATAC AGCGTGTCCT TGCGGCGGCT GCAAGAGGAC
151 TAAGCATGGA TGGCAGCCGG AGAGTCAGAG CAACCTCTGT CCTTCCAGAG
201 TATGGTCCAC CGTGCCATT TAAAGGACAC TTGAGCACC AAGTAATGC
251 TGCAGTAGAC TGCTCGGTTC CAGTAAGCAT GAGTACCAGC ATAAAGTATG
301 CAGACCAACA ACGAAGAGAG AAATCAAAA AGGAATTAGC ACAATGTGAA
351 AAAGAGTTCA AATTAATAA AACTGCAATG CGAGCCAATT ATAAAAATAA
401 TTCCAAGTCA CTTTTTAATA CCTTACAAGA GCCCTCAGGC GAACCGCAAA
451 TTGAGGATGA CATGTTAAAA GAAGAAATGA ATGGATTTC ATCCTTTGCA
501 AGGTCACTAG TACCCTCTTC AGAGAGACTA CACCTAAGTC TACATAAATC
551 CAGTAAAGTC ATCACAATG GTCTGAGAA GAACTCCAGT TCCTCCCGCT
601 CCAGTGTGGA TTATGCAGCC TCCGGGCCCC GAAACTGAG CTCTGGAGCC
651 CTGTATGGCA GAAGGCCAG AAGCACATTC CCAAAATCCC ACCGGTTTCA
701 GTTAGTCATT TCGAAAGCAC CCAGTGGGGA TCTTTGGAT AAACATTCTG
751 AACTCTTTTC TAACAAACAA TTGCCATTCA CTCCTCGCAC TTTAAAAACA
801 GAAGCAAAAT CTTTCTGTG ACAGTATCGC TATTATACAC CTGCCAAAAG
851 AAAAAAGGAT TTTACAGATC AACGGATAGA AGCTGAAACC CAGACTGAAT
901 TAAGCTTTAA ATCTGAGTTG GGGACAGCTG AGACTAAAAA CATGACAGAT
951 TCAGAAATGA ACATAAAGCA GGCATCTAAT TGTGTGACAT ATGATGCCAA
1001 AGAAAAAATA GCTCCTTTAC CTTTAGAAGG GCATGACTCA ACATGGGATG
1051 AGATTTAAGGA TGATGCTCTT CAGCATTCTT CACCAAGGGC AATGTGTCAG
1101 TATTCCTGTA AGCCCCCTTC AACTCGTAAA ATCTACTCTG ATGAAGAAGA
1151 ACTGTTGTAT CTGAGTTTCA TTGAAGATGT AACAGATGAA ATTTTGAAC
1201 TTGGTTTATT TTCAAAACAGG TTTTAGAAG GACTGTTTCA GCGACATATA
1251 AAACAAATA AACATTGGA GGGGAAAAA ATGCGCCACC TGCTGCATGT
1301 CCTGAAAGTA GACTTAGGCT GCACATCGGA GAAACTCG GTAAAGCAAA
1351 ATGATGTTGA TATGTTGAAT GTATTTGATT TTGAAAAGGC TGGGAATTCA
1401 GAACCAATA AATTAATAA TGAAGTGAA GTAACAATTC AGCAGGAACG
1451 TCAACAATAC CAAAAGGCTT TGGATATGTT ATTGTCGCA CCAAGGATG
1501 AGAACGAGAT ATTCCCTTCA CCAACTGAAT TTTTCATGCC TATTATAAA
1551 TCAAGCATT CAGAAGGGGT TATAATTCAA CAGGTGAATG ATGAAACAAA
1601 TCTTGAACAT TCAACTTTGG ATGAAATCA TCCAAGTATT TCAGACAGTT
1651 TAACAGATCG GAAACTTCT GTGAATGTCA TTGAAGTGA TAGTGACCTT
1701 GAAAAGGTTG AGATTTCAA TGGATTATGT GGTCTTAACA CATCACCTTC
1751 CCAATCTGTT CAGTTCTCCA GTGTCAAAGG CGACAATAAT CATGACATGG
1801 AGTTATCAAC TCTTAAATC ATGGAATGA GCATTGAGGA CTGCCCTTTG
1851 GATGTTTAA CTTCATTAAT AAATACCTCA AATGGCCAGT AAAAAAATAA
1901 AAAAAAATAA
```

BLAST Results

Entry HS477360 from database EMBL:

human STS WI-14643.

Length = 418

Minus Strand HSPs:

Score = 1850 (277.6 bits), Expect = 2.5e-77, P = 2.5e-77

Identities = 392/405 (96%), Positives = 392/405 (96%), Strand = Minus / Plus

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 156 bp to 1856 bp; peptide length: 567
 Category: similarity to unknown protein

```

1 MDGSRVRAT SVLPYRGPPC LKGLHSTKS NAAVDCSVPV SMSTSIKYAD
51 QQRREKLKKE LAQCEKEFKL TKTAMRANYK NNSKSLFNTL QEPGEPQIE
101 DDMLKEEMNG FSSFARSLVP SSERLHLSLH KSSKVITNGP EKNSSSSPSS
151 VDYAASGPRK LSSGALYGRR PRSTFPNSHR FQLVSKAPS GDLLDKHSEL
201 FSNKQLPFTP RTLKTEAKSF LSQYRYTPA KRKKDFTDQR IEAETQTELS
251 FKSELGTAET KNMTDSEMNI KQASNCVTYD AKEKIAPLPL EGHDSWDEI
301 KDDALQHSSP RAMCQYSLKP PSTRKIYSE EELLYLSFIE DVTDEILKLG
351 LFSNRFLERL FERHIKQNKH LEGEKMRHLL HVLKVDLGCT SEENSVKQND
401 VDMLNVDFDE KAGNSEPNKL KNESEVTIQQ ERQQYQKALD MLLSAPKDN
451 EIFPSPTEFF MPIYKSKHSE GVIIQQVNDE TNLETSTLDE NHPSISDSLT
501 DRETSVNVEI GDSOPEKVEI SNGLCGLNTS PSQSVQFSSV KGDNNHDMEL
551 STLKIMEMSI EDCPLDV

```

BLASTP hits

Entry CEC18C4_3 from database TREMBL:
 "C18C4.5"; Caenorhabditis elegans cosmid C18C4.
 Length = 1091
 Score = 98 (34.5 bits), Expect = 0.29, P = 0.25
 Identities = 105/470 (22%), Positives = 192/470 (40%)

Alert BLASTP hits for DKFZphfbr2_22f21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_22f21, frame 3

Report for DKFZphfbr2_22f21.3

```

[LENGTH]      567
[MW]           64120.02
[pi]           5.68
[PROSITE]      AMIDATION      1
[PROSITE]      MYRISTYL       3
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 16
[PROSITE]      PKC_PHOSPHO_SITE 18
[PROSITE]      ASN_GLYCOSYLATION 4
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 1.23 %

```

```

SEQ MDGSRVRATSVLPYRGPPCLFKGLHSTKSNAAVDCSVPVSMSTSIKYADQQRREKLKKE
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ LAQCEKEFKLTKTAMRANYKNNKSLFNTLQEPGEPQIEDMLKEEMNGFSSFARSLVP
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ SSERLHLSLHKSSKVITNGPEKNSSSSPSSVDYAASGPRKLSSGALYGRRPRSTFPNSHR
SEG .....
PRD cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ FQLVSKAPSGDLLDKHSELSFNKQLPFTPRTLKTEAKSFLSQYRYTPAKRKKDFTDQR
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ IEAETQTELSFKSELGTAETKNMTDSEMNIKQASNCVTYDAKEKIAPLPLEGHDSWDEI
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```



```

SEQ      KDDALQHSSPRAMCQYSLKPPSTRKIYSDEEELLYLSFIEDVTDEILKLGLFSNRFLERL
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      FERHIKQNKHLEGEKMRHLLHVLKVDLGCTSEENSVKQNDVMDLNVDFEKGAGNEPNKL
SEG      .....
PRD      hhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      KNESEVTIQERQQYQKALDMLLSAPKDENEIFPSPTEFFMPIYKSKHSEGVIIQQVNDE
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      TNLETSLDENHPSISDSLTDRETSVNVIEGDSDPKVEISNGLCGLNTSPSQSVQFSSV
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      KGDNNHDMELSTLKIMEMSIEDCPLDV
SEG      .....
PRD      cccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

Prosites for DKFZphfbr2_22f21.3

PS00001	81->85	ASN_GLYCOSYLATION	PDOC00001
PS00001	143->147	ASN_GLYCOSYLATION	PDOC00001
PS00001	262->266	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00004	159->163	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00005	178->181	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	212->215	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	309->312	PKC_PHOSPHO_SITE	PDOC00005
PS00005	317->320	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	353->356	PKC_PHOSPHO_SITE	PDOC00005
PS00005	395->398	PKC_PHOSPHO_SITE	PDOC00005
PS00005	500->503	PKC_PHOSPHO_SITE	PDOC00005
PS00005	539->542	PKC_PHOSPHO_SITE	PDOC00005
PS00005	552->555	PKC_PHOSPHO_SITE	PDOC00005
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	264->268	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	337->341	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	481->485	CK2_PHOSPHO_SITE	PDOC00006
PS00006	486->490	CK2_PHOSPHO_SITE	PDOC00006
PS00006	494->498	CK2_PHOSPHO_SITE	PDOC00006
PS00006	498->502	CK2_PHOSPHO_SITE	PDOC00006
PS00006	500->504	CK2_PHOSPHO_SITE	PDOC00006
PS00006	513->517	CK2_PHOSPHO_SITE	PDOC00006
PS00006	559->563	CK2_PHOSPHO_SITE	PDOC00006
PS00008	164->170	MYRISTYL	PDOC00008
PS00008	256->262	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00009	167->171	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_22f21.3)

DKFZphfbr2_22h13

group: transmembrane protein

DKFZphfbr2_22h13 encodes a novel 520 amino acid protein, with similarity to *Drosophila melanogaster* EG:39E1.3.

The protein contains an ATP/GTP A Prosite pattern (P-loop). This loop interacts with one of the phosphate groups of a A or G nucleotide. It is found in numerous ATP- or GTP-binding proteins, such as ATP synthase alpha and beta subunits, Myosin heavy chains, Kinesin heavy chains and kinesin-like proteins, Dynamins and dynamin-like proteins, several kinases, DNA and RNA helicases, GTP-binding elongation factors and the Ras family of GTP-binding proteins. Additionally, the novel protein contains one putative transmembran domain.

No informative BLAST results; no predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

AC004780_1, differences to predicted genmodel

membrane regions: 1

AC004780_1, differences to predicted genmodel

complete cDNA, complete cds, EST hits
on genomic level encoded by AC004780,
differences to predicted genmodel!
TRANSMEMBRANE 1

Sequenced by AGOWA

Locus: unknown

Insert length: 2292 bp

Poly A stretch at pos. 2272, polyadenylation signal at pos. 2255

```
1 GGGGGAGGGA ACTGATCTCA GCTCGGGCCC GCGTTACATC CTCCTCCTCT
51 TCTTCCTTCG GCCCAGCTTT CCTTAGGGGC TGCAACCCGG ACGCCGAGGC
101 CGGTTTCGGA GTGGGGAGTG CCCATTTTCT CTCTTCCCA CGTTCTCTGGC
151 CCCAGACGC CATTTCAGG CGGGTGGCTT GGGTCAGCCT CCCC GCCCCC
201 ACCGACTCC CGTCACGGGA GAGCGCACAC CGCGCCCCGA GAACCAATCA
251 GCAGCCCGGT TAGGTAACCA TGTCTGAGTC TGGACACAGT CAGCCTGGAC
301 TCTATGGGAT AGAGCGGCGG CGACGGTGGG AGGAGCCTGG CTCTGGTGGC
351 CCCAGAATC TCTCTGGGCC TGGTGGTGGG GAGAGGGACT ACATTGCACC
401 ATGGGAAAGA GAGAGAAGGG ATGCCAGCGA AGAGACAAGC ACTTCCGTCA
451 TGCAGAAAC CCCCATCATC CTCTCAAAAC CTCCAGCAGA GCGGTCAAAA
501 CAGCCACCA CTTCAACAGC CCCTGCTGCC CCGCTGCTC CAGCCCTCT
551 GGAGAAGCCC ATCGTTCTCA TGAAGCCACG GGAGGAGGGG AAGGGGCTG
601 TGGCCGTGAC AGGTGCCTCT ACCCCTGAGG GCACCGCCCC ACCACCCCT
651 GCAGCCCTG CGCCACCCAA GGGGGAGAAG GAGGGGCAGA GACCCACACA
701 GCCTGTGTAC CAGATCCAGA ACCGGGGCAT GGGCACTGCC GCACCAGCAG
751 CCATGGACCC TGTCGTGGGT CAGGCCAAAC TACTGCCCCC AGAGCGCATG
801 AAGCACAGCA TCAAGTTGGT GGATGACCAG ATGAATTGGT GTGACAGTGC
851 CATCGAGTAC CTGTTGGATC AGACTGATGT GTTGGTGGTT GGTGTCCTGG
901 GCCTCCAGGG GACAGGCAAG TCCATGGTCA TGTCATTGTT GTCAGCCAAC
951 ACTCCAGAGG AGGACCAGAG GACTTATGTT TTCCGGGCCC AGAGCGCTGA
1001 AATGAAGGAA CGAGGGGGCA ACCAGACCAG TGGCATCGAC TTCTTTATTA
1051 CCCAAGAACG GATTGTTTTC CTGGACACAC AGCCCATCCT GAGCCCTTCT
1101 ATCTAGACC ATCTCATCAA TAATGACCGC AAATGCGCTC CAGAGTACAA
1151 CCTTCCCAC ACTTACGTTG AAATGCAGTC ACTCCAGATT GCTGCCTTCC
1201 TTTTCACGGT CTGCCATGTG GTGATTGTTG TCCAGGACTG GTTCACAGAC
1251 CTCAGTCTCT ACAGGTTCCCT GCAGACAGCA GAGATGGTGA AGCCCTCCAC
1301 CCCATCCCCC AGCCACGAGT CCAGCAGCTC ATCGGGCTCC GATGAAGGCA
1351 CCGAGTACTA CCCCACCTA GTCTTCTTGC AGAACAAGC TCGCCGAGAG
1401 GACTTCTGTC CTCGGAAGCT GCGGCAGATG CACCTGATGA TTGACCAGCT
1451 CATGGCCAC TCCCACCTGC GTTACAAGGG AACTCTGTCC ATGTTACAAT
1501 GCAATGTCTT CCGGGGGCTT CCACCTGACT TCCTGGACTC TGAGGTCAAC
1551 TTATTCCTGG TACCCTTCAT GGACAGTGAA GCAGAGAGTG AAAACCCACC
1601 AAGAGCAGGA CCTGGTTCCA GCCCACTCTT CTCCCTGCTG CTTGGGTATC
1651 GTGGCCACCC CAGTTTCCAG TCCTTGGTGA GCAAGCTCCG GAGCCAAGTG
1701 ATGTCCATGG CCGGCCACA GCTGTCACAC ACGATCCTCA CCGAGAAGAA
1751 CTGGTTCCAC TACGCTGCCG GATCTGGGA TGGGGTGAGA AAGTCCTCTG
1801 CTCTGGCAGA GTACAGCCGC CTGCTGGCCT GAGGCCAAGG AGAGGAATGT
1851 CATGCAGGGG ACCTCCTGGG TCCGAGTGT ACTGCGAGGG AGCAGAGATG
1901 TCCATCCCCC GCTGGGGTGG AGAGCGGCAG CAGGCCTGAT GGATGAGGGA
1951 TCGTGGCTTC CCGGCCAGA GACATGAGGT GTCCAGGGCC AGGCCCCCA
```

```

2001 CCCTCAGTTG GGGCTGTTCC GGGGGTGA CTGTGAGCGATC CCACCCCAAA
2051 CCTGAGATGG GGTAGCCCGT CCTGTGTCCT CCACAGGGAC AAGCAGTGGG
2101 AGGAGTCTGA ATGGTCACCA GGAAGCCCGG GCTCCATCTT GACCTCCTTT
2151 TTCAGGGACA GGAGCAACAG GCCCCTCTTC CCTGACTCTA AGCCCTTCCC
2201 TGTAAGGTGA GGCAGGGTCT GGAGAGCTCT TTATTGGAAC AGATCTGGTG
2251 GTTCAAATAA ACACAGTCAT GCAAAAAAAA AAAAAAAA AA

```

BLAST Results

Entry AC004780 from database EMBL:
Homo sapiens chromosome 19, cosmid F17127, complete sequence.
Score = 2616, P = 0.0e+00, identities = 524/525
15 exons Bp 8031-31789

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 270 bp to 1829 bp; peptide length: 520
Category: similarity to unknown protein
Prosite motifs: ATP_GTP_A (211-219)

```

1 MSESQHSQPG LYGIERRRRW KEPGSGGPQN LSGPGGRERD YIAPWERERR
51 DASEETSTSV MQKTPIIILSK PPAERSKQPP PPTAPAAPPA PABLEKPIVL
101 MKPREEGKGP VAVTGASTPE GTAPPPPAAP APPKGEKEGQ RPTQPVYQIQ
151 NRGMGTAAPA AMDPVVGQAK LLPPERMKHS IKLVDDQMNW CDSAIEYLLD
201 QTDVLVVGVL GLQGTGKSMV MSLLSANTPE EDQRTYVFRA QSAEMKERGG
251 NQTSIDFFFI TQERIVFLDT QPILSPSILD HLINNDKRLP PEYNLPHTYV
301 EMQSLQIAAF LFTVCHVVIV VQDWFDTLSL YRFLQTAEMV KPSTPSPSHE
351 SSSSSGSDG TEYYPHLVFL QNKARREDFC PRKLQMHLM IDQLMAHSHL
401 RYKGTLSMLQ CNVFPGLPPD FLDSEVNLFV VPFMDSEAES ENPPRAGPGS
451 SPLFSLPGY RGHPFSQSLV SKLRSQVMSM ARPQLSHTIL TEKNWFHYAA
501 RIWDGVRKSS ALAEYSRLLA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_22h13, frame 3

TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19,
cosmid F17127, complete sequence., N = 2, Score = 1264, P = 1.3e-231

TREMBL:CEY54E2A_1 gene: "Y54E2A.2"; Caenorhabditis elegans cosmid
Y54E2A, N = 2, Score = 219, P = 1.4e-15

>TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid
F17127, complete sequence.
Length = 528

HSPs:

Score = 1264 (189.6 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231
Identities = 254/302 (84%), Positives = 264/302 (87%)

```

Query: 46 ERERRDASEETSTSVMQKTPIIILSKPPAERSKQPPPTAPAAPPAPAPLEKPIVLMKPRE 105
      E+ER D+ + S +Q+T + R + P + A APLEKPIVLMKPRE
Sbjct: 39 EKER-DSDSDFSP--LQTEGCRDRKHFRHAENPHHPLKTSSRA-APLEKPIVLMKPRE 94

Query: 106 EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAAMPDV 165
      EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAAMPDV
Sbjct: 95 EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAAMPDV 154

Query: 166 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 225
      VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS
Sbjct: 155 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 214

```

Query: 226 ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN 285
 ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN
 Sbjct: 215 ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN 274

Query: 286 DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFDTLSLYRFLQTAEMVKPSTP 345
 DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFDTLSLYR K ++
 Sbjct: 275 DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFDTLSLYRLWDLGCKCKSNSH 334

Query: 346 SP 347
 SP
 Sbjct: 335 SP 336

Score = 993 (149.0 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231
 Identities = 189/189 (100%), Positives = 189/189 (100%)

Query: 332 RFLQTAEMVKPSTPSPSHESSSSSSGSDEGTEYYPHLVFLQNKARREDFCPRKLRQMHLMI 391
 RFLQTAEMVKPSTPSPSHESSSSSSGSDEGTEYYPHLVFLQNKARREDFCPRKLRQMHLMI
 Sbjct: 340 RFLQTAEMVKPSTPSPSHESSSSSSGSDEGTEYYPHLVFLQNKARREDFCPRKLRQMHLMI 399

Query: 392 DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS 451
 DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS
 Sbjct: 400 DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS 459

Query: 452 PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPOLSHITLTKNWFHYAARIWDGVRKSSA 511
 PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPOLSHITLTKNWFHYAARIWDGVRKSSA
 Sbjct: 460 PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPOLSHITLTKNWFHYAARIWDGVRKSSA 519

Query: 512 LAEYSRLLA 520
 LAEYSRLLA
 Sbjct: 520 LAEYSRLLA 528

Pedant information for DKFZphfbr2_22h13, frame 3

Report for DKFZphfbr2_22h13.3

[LENGTH] 520
 [MW] 57650.81
 [pI] 6.52
 [HOMOL] TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid
 F17127, complete sequence. 0.0
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 8
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 8
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 3
 [PROSITE] ASN_GLYCOSYLATION 2
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 11.73 %

SEQ MSEGHSQPGLYGIERRRRWKEPGSGGPQNLSGPGRERDYIAPWERERRDASEETSTSV
 SEG
 PRD ccc
 MEM

SEQ MQKTPILSKPPAERSKQPPPTAPAAPAPAPLEKPIVLMKPREEGKGPVAVTGASTPE
 SEGxxxxxxxxxxxxxx
 PRD eecceeecc
 MEM

SEQ GTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAMDVPVQAKLLPPERMKHS
 SEG ..xxxxxxxxxxxxxx
 PRD ccc
 MEM

SEQ IKLVDDQMNWCDIAEYLLDQTDVLVVGVLGLQGTGKSMVSLLSANTPEEDQRTYVFRA
 SEGxxxxxxxxxxxxxxxxxxxxxx
 PRD hhhhhcc
 MEM

SEQ QSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINNDKLPPEYNLPHTYV
 SEG
 PRD hhhhhhccchh
 MEM

SEQ EMQSLQIAAFLFTVCHVVIVVQDWFDTLSLYRFLQTAEMVKPSTPSPSHESSSSSSGSDEG
 SEGxxxxxxxxxxxxxxxxxxxxxx

```

PRD      hhhhhhhhhhhhhhhheeeeeccchhhhhhhhhhhhhcccccccccccccccccc
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
SEQ      TEYYPHLVFLQNKARREDFCPRKLRQMHLMDQLMAHSHLRYKGTLSMLQCNVFPGLPPD
SEG      .....
PRD      cccccceeehhhhhhccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc
MEM      .....
SEQ      FLDSEVNLFLVPFMDSEAESENPPRAGPGSSPLFSLPGYRGHPSFQSLVSKLRSQVMSM
SEG      .....
PRD      chhhhhhheeeeeccccccccccccccccccccceccccccccchhhhhhhhhhhhhh
MEM      .....
SEQ      ARPQLSHTILTEKNWFHYAARIWDGVRKSSALA EYSRLLA
SEG      .....
PRD      hhhhhhhhheeeccchhhhhhhhhhhhhcchhhhhhhhhhhcc
MEM      .....

```

Prosites for DKFzphfbr2_22h13.3

PS00001	30->34	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00002	32->36	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	507->511	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	215->218	PKC_PHOSPHO_SITE	PDOC00005
PS00005	491->494	PKC_PHOSPHO_SITE	PDOC00005
PS00006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	254->258	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	298->302	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	436->440	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	211->217	MYRISTYL	PDOC00008
PS00008	214->220	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	356->362	MYRISTYL	PDOC00008
PS00008	505->511	MYRISTYL	PDOC00008
PS00017	211->219	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFzphfbr2_22h13.3)

DKFZphfbr2_22i4

group: brain derived

DKFZphfbr2_22i4.1 encodes a novel 228 amino acid protein with similarity to the N-terminus of human p52rIPK.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Human P52rIPK N-terminus

complete cDNA, complete cds, few EST hits
function of P52rIPK, repressor of p58IPK protein kinase inhibitor
upstream regulator of interferon induced proteins

Sequenced by AGOWA

Locus: unknown

Insert length: 4748 bp

Poly A stretch at pos. 4726, polyadenylation signal at pos. 4709

```
1 TGGGTCCGGT CCTAGGGTCA CACCCACCGC AGGGTCTGGC TTGGTACAGT
51 TGGGTGCATG CAGAAGTAGG TGGAGCTGCT GTTGCAGCCT TGAGAGAGTT
101 TTATTGTAAA ACTCTTGTA TTTATAGTAA TCGGAGGGGA AAACACCTCT
151 TCCTTTTAAT TGCTCTGAGG ACCGCTGCCA AAGAAACGCA GTAGATCCGC
201 TCCCTCTTGG GGGCGGGGAG AAAGAACGGG TTGTGTCCGC CATGTTGGTG
251 AAGTCAAGCG AAGGCGACTA GAGCTCCAGG AGGGCCAGTT CTGTGGGCTC
301 TAGTCGGCCA TATTAATAAA GAGAAAGGGA AGGCTGACCG TCCTTCGCCT
351 CCGCCCCCAC ATACACACCC CTTCTTCCCA CTCGCTCTC ACGACTAAGC
401 TCTCACGATT AAGGCACGCC TGCCTCGATT GTCCAGCCTC TGCCAGAAGA
451 AAGCTTAGCA GCCAGCGCCT CAGTAGAGAC CTAAGGGCGC TGAATGAGTG
501 GGAAGGGGAA ATGCCGACCA ATTGCGCTGC GCGGGGCTGT GCCACTACCT
551 ACAACAAGCA CATTAACATC AGCTTCCACA GGTTCCTTT GGATCCTAAA
601 AGAAGAAAAG AATGGGTTTCG CCTGGTTAGG CGCAAAAATT TTGTGCCAGG
651 AAAACACACT TTTCTTTGTT CAAAGCACTT TGAAGCCTCC TGTTTGACC
701 TAACAGGACA AACTCGACGA CTTAAATGG ATGCTGTTC AACCATTTTT
751 GATTTTTGTA CCCATATAAA GTCTATGAAA CTCAAGTCAA GGAATCTTTT
801 GAAGAAAAAC AACAGTTGTT CTCCAGCTGG ACCATCTAAT TTAATCAAA
851 ACATTTAGTAG TCAGCAAGTA CTACTTGAAC ACAGCTATGC CTTTAGGAAT
901 CCTATGGAGG CAAAAAAGAG GATCATTAAT CTGGAAGGAA AAATAGCAAG
951 CTTAAGAAGA AAAATGAAAA CTTGCCTACA AAAGGAACGC AGAGCAACTC
1001 GAAGATGGAT CAAAGCCACG TGTGTTGTA AGAATTTAGA AGCAAAATAGT
1051 GTATTACCTA AAGGTACATC AGAACACATG TTACCAACTG CCTTAAGCAG
1101 TCTTCCCTTG GAAGATTTTA AGATCCTTGA ACAAGATCAA CAAGATAAAA
1151 CACTGCTAAG TCTAAATCTA AAACAGACCA AGAGTACCTT CATTTAAATT
1201 TAGCTTGCAC AGAGCTTGAT GCCTATCCTT CATTCCTTTC AGAAGTAAAG
1251 ATAATTATGG CACTTATGCC AAAATTCATT ATTTAATAAA GTTTTACTTG
1301 AAGTAAACAT ACTGAATTTG TGAAGACTTG ATTACAAAAG AATAAAAAAC
1351 TTCATATGGA AATTTTATTT GAAAATGAGT GGAAGTGCTT TACATTAGAA
1401 TTACGGACTT AAAAAATTTG CTAATAAATT GTGTGTTGA AAGGTGTTTT
1451 TTGTTTTTGT CTTTTTAAAC TACTGTAAAA AGAACAGCTT ATGATAAGTA
1501 ATATGTTTAA CTTAGAGAAG AATTTTTTCC TGTACCAAAG TTGGCATATT
1551 GCATTCTAAA TAAGATGCTA AATAAGAGTT AACCAACATT CAACATGACC
1601 TTAAGACTGC TGGGTTTTGT ATTAATTAAT TTATAATTGG CACTGTGATT
1651 TGAAAAATTT ATAGAAAAAA AGGTACAGGG CAAGTTTTTA AATTAAGACT
1701 TTCTATATTT TGTGTTTACCA GTAAAAGTGA GCTTATCATG GCCTCTCTCA
1751 TAAGAATGAT TTTAAATAG GTTGTAAAT ATTTTGAAAA TATTTGAATG
1801 TGAAGTACCA TTGAGTCATC CAACTAGGTT AAGGCCCTCA GTACTTTAAA
1851 CTAGTAAAAA CTAGTAGCTG ATAATATTCA CCTAAGTAAG TGTGTAAAAA
1901 TAATTCAGAG TTCAGGACCT AGCTTAGATA AATGTATACT ACTCTTTTTC
1951 TCATAGTAAA AATCTTACAT TTCCAACCTC AAAATTGGTG CTTCATATT
2001 TGTGTAATAA CAAAACCTCT AAGGTTTTTT GTTTCTTTT TAACACTTTT
2051 CCAAAATGCAT ACTATACCTC AGAAATAGTG TATCAATATA GTGGGCTTTT
2101 TTTTTCCTCT TCATAAACC ACAGTAAAT TTAATCACAG GAACTACTT
2151 ATATCTTTC ACCTTTGTATT GATAACTTAA AATGGCATCA GTTTATCTTA
2201 GACATCAGCT TGCTTTTAT CTCCTTTTTT AGTGAGTGAA ATAGAGCAAC
2251 TAGCATGCCT GTGTTCCAG CTACTTGGGA GGCTAAGGTG GGAAGATCAA
2301 TTGAACCTAG GAGGTTGAGG CTATAGTGAG CTGTGATTGC ACGACTGCAC
2351 TCAGGCTCGG GCAATGGAGT GAGACTCCTG TCTCTAAAC AGCAACAACA
2401 AAAATAAAGC AACCATAGTG CATAAGGGAA ATTAATGTT CCTATAGAA
2451 ATATGTGTAT GTCTGTGATA GTGGTATGCA AATGCTAATT ATTTTATAAA
2501 ATAAAAGTTC AGAATATTTC TTATCATTCG CACTTGAACA ATTAAAGGGT
2551 TTGCTTTATT TCACTAATGT TTAATAGGAA CCCTTTCCTT CAAACAGCTT
```

```

2601 TGTGAAATC ATGTAAAAAT TTGTTAATAG AGAATCAAGT TATTTAACTC
2651 AACTTATTTA ATTCAAGCTT GTGATACTAA CATACAAAGG TAGCATAAAC
2701 CAAGTCATAA ATTGCTGTAA TCTTTCCTGT AGAGTAATAG CTACTTCATG
2751 ATTTTAAAAA AAATTTTCATT TTTTGTCTAT TTAGGATTGC ATTTGCTTGG
2801 CTCCTAGTAA CAATTCTTTT ACAGTATTAG CACTCTCTTT ACTAAGGAAT
2851 GCCTCCCAAG GAAATGCAAA GGTAGGAAAA GTCTCTTAGA ATGCCCATGA
2901 GGTATTTAAA ACAGATATTT ATGAAAAATCT TTTTGTGAAT GTTATAAATC
2951 TTGCTAGTTA TTTTATCTTT ATCTTAAGTA TTAGATGTAG TTCCTTGGAA
3001 TTGTCATTAC ATATTTATTT TTTTCTAGTG TGGTTTCAA TAACTTTTTG
3051 CCAACATATA ATCATCATCA AACATTCACT GACCATATCT ATTTTATAAC
3101 TCAAAATAAG TTGGACAAAT AATCATTTTA ATAAAACTA TTTTTCCTAA
3151 GTATAACCAAC TGTCATGTGG TTCACCCCTC ACCCCAGATA CAAAACACTT
3201 ATTTGTGTAG CCCAGTTCCT ATCTACAGTA ATACCTTGAA ACCTTAATAA
3251 ATTTTAAAAA TCATAAAAAA AAAATATTGT AAAATACAAC AAATTTTGGG
3301 CAAGGTTACT TCATCTTCAT TCATTATTAC CTGACAGTAT TAAACTACTA
3351 CTCATTAATT TTAGAGTAAA CTTTCTGTG TTTTCCCGT GATTTTCATT
3401 GTGCTGTCCT GACAACATGC TCCAAACTCT TTGCATCAA TTGTTTTATT
3451 AACATACATT TGCTACCTT AAAACTAGCT TTATTCACAG AGAAAGACCT
3501 AAAAGGAGTC TATTAATATG CTGCTTTCAG TTTGATAGTT TTTTAAAAA
3551 TCACTCTGAC CATAAACTAA CTGAAATAT ATATGGATTT TTTTCTCTC
3601 CCGGTCACAA CACAGATCTT CTGTTTCTT GTTCTCTGTC TACTGGGCAC
3651 CAACCTCTAC AAAGAACCAG CCAAAGGCTA GGTACTTGAT ATAAAAAGGA
3701 ATATTACATT ATTTTCTGCC CTCAAGTTGC TCTATCTCCT GAAAGAAACA
3751 AGTAATATTT ATAATACAAT ATGATAAATG CTACAAAAGA AATAGCTGTA
3801 AAGTCCCTTG GTAAATGCTG TTGAATTGGA ATTCAAGTAAG AACTATAAAC
3851 TGTAGACCTT TTTATAATCA AATGCTTTTG TCTTGAAACA AAACAGATTC
3901 CTCCTTATAT TGACTTAGCA AAGGAGGTAC AAGGACATTG GCATTTGACC
3951 TGAATTATGG TGTTTTATTG AATGAGCTAT AAGACAACAT TTTTACCCTT
4001 TAAATGAAC ACTGAACAAA TGTGTTAATG GTATCTTTGT TAAAAGGAAA
4051 ACATAGCTAT AAATAAAATA CTACATCGAA ATCCAGCACT GGAGTTCATT
4101 TGAAATTTGA TATTTTGTGT AAAGTAACAA ACCTATTAA ACAGATTTT
4151 AAAATAAATC AGAATCGTAT AAAGCACTTT GGTACTTATT TGTTCTCTTT
4201 TCCCTTACAT TCTGTGTGGT AGGTGGTATT ATCTCTGATT TACACATGAA
4251 GACATCCTTG TTAATGCAAT TTATTTATTC ATTCGGGCAT TTACTGTGTG
4301 CCAACTTGCA AAAGGAATAG AAATGTCTGT GATCTAGATA GTTCTAGATT
4351 GAACATAGAT TTTCTGCCAA CAAATCCTCT CTGCTGTTCA CATTATCCTT
4401 TGTTTAACGT ATGAACCAGG TTAATAAAAT AGGATAAATC ATGTGTCTTA
4451 GAATATGAAA ATAGTAAGGT CTTTGAGGTC ACTTGATCTT CTCTAAGTAG
4501 ACTTTATAAT ATTGTGTTTT ATCTCATTTT TCAATATTAG AATACGGGTA
4551 GATTTTAATT TTGCTATAAT ATAGGAAATG GTTCATCTTT GTACCAAAAT
4601 ATTGCATTCT TCTGATATTT AGACAGTTGG AAACCTTCTA AAATTGAGGA
4651 TTTTGTAGTG TATACTAAAT AATTGCATAT TCAAAAAAAT GTATTCTGAG
4701 TATGGTGATA TTAACATTTT TTCCCCAAAA AAAAAAATAA AAAAAAAT

```

BLAST Results

No BLAST result

Medline entries

98107671:
Regulation of interferon-induced protein kinase PKR:
modulation of P58IPK inhibitory function by a novel protein,
P52rIPK

Peptide information for frame 1

ORF from 511 bp to 1194 bp; peptide length: 228
Category: similarity to known protein

```

1 MPTNCAAAGC ATTYNKHINI SFHRFPLDPK RRKEWVRLVR RKNFVPGKHT
51 FLCSKHFEAS CFDLTGQTRR LKMDAVPTIF DFCTHIKSMK LKSRNLLKKN
101 NSCSPAGPSN LKSNISSQOV LLEHSYAFRN PMEAKKRIK LEKEIASLRR
151 KMKTCLOKER RATRRWIKAT CLVKNEANS VLPKGTSEHM LPTALSSSLPL
201 EDFKILEQDQ QDKTLLSLNL KQTKSTFI

```

BLASTP hits

Entry AF007393_1 from database TREMBL:
product: "P52rIPK"; Homo sapiens P52rIPK mRNA, complete cds.
Score = 166, P = 2.5e-11, identities = 40/106, positives = 56/106

DKF2phfbr2_22k3

group: brain derived

DKF2phfbr2_22k3 encodes a novel 538 amino acid protein with weak similarity to extensins.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to extensins

complete cDNA, complete cds, few EST hits
CpG Island in 5' UTR complete cDNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2775 bp

Poly A stretch at pos. 2755, polyadenylation signal at pos. 2718

```
1 GGGGCTGCCC GCGCGCTCCA CGGTGCAGAG CTCTAAGCGC GCGGGCTGGC
51 AGGCTGCGGC GCGTCAAGGT CAGCCTGGAG CTGGGTGGCG GCCTGCCTGG
101 GGGCGGGGGA CCCTACTGGA GGCCCGGCT GGGCCTCCC AGCGCCTCGG
151 CCATATTGAA TAGCTTCGAC TGGACCGTCT TTGTCTGCGA AGTCCTGTCC
201 CAAGTTCCAG CCGCGTCCCT GGGGCTGGG GCAGGAAGAG TCGCTGGCAG
251 CCGCGCGGCC CCAACTTGA GCTGGGACAC CACGTTTCCA GCTTGGAGTG
301 GGCCTTGAGC CTTGGGACTG ACCTCGCCCC CGGCTCAGT AGGCATCCTG
351 GAAATTGATT CCCCCAAGTC CTTGGTGGG GAGCCGACT TGGTCAAGAC
401 TGTACTTGTT GCAGGCGAAG AGATTGGAGG CGTTTGGCTC GTCCCTGGCT
451 AGGGAGGTGA GACTCTCCGG TCAGCGTTGC TGAACCTCCC CCCATCCAGT
501 CCCTCCCTCA AGACTAAGGG CTACAGTAGT TTGTTGGGGC TCATTGCCCC
551 CTCACCCAG ATATCACCTT GGAGATCTTA AAGACTCTCG AGAAAAGCCA
601 CGTGGGGGGC TGGTTCCTCT GGGGCTTCT GCGTCCCCC GACTGCCTCA
651 TTCTTTGGAG CGTCCCGAT GTCTGCAAG ATGTGGATT GGACGTCCTC
701 GTGGAAGCCC TAAAGCCCGT GGGGACATTT AAGAAGATCG GCAAGGTGTT
751 CCGCAAGGAG GAGGACTCCA CGGTGGGGAT GCTGCAGATC GGGGAGGACG
801 TCGACTATTT GCTCATCCCC CGGGAGGTCA GGCTGGCTGG GGCCTCTGG
851 AGAGTCATCT CTAAGCCCGC CACCAAGGAA GCAGAATTC GGGAGCGGCT
901 GACCCAGTTC CTGGAAGAAG AGGGCCGCAC CTTGGAGGAC GTGGCCCGCA
951 TCATGGAGAA GAGCACCCCG CACCCGCCCC AGCCCCCAA AAAGCCCAAG
1001 GAGCCCCGAG TGAGGAGGAG AGTGCAGCAG ATGGTGACTC CTCGCCCCCG
1051 GCTGGTCTGT GGCACGTACG ACAGCAGCAA CGCCAGCGAC AGCGAGTTCA
1101 GCGACTTCGA GACCTCCAGA GACAAGAGCC GCCAGGGCCC GCGCGGGGGC
1151 AAGAAGGTGC CAAAATGCC CGTCAGCTAC CTGGGCAGCA AGTTCCTGGG
1201 AAGCGACCTG GAGAGTGAGG ATGATGAGGA ACTGGTCTGAG GCCTTCCTCC
1251 GCGGCAGGAG GAAGCAGCCC AGCGCGCCCG CTGCCCCGCG CCGCGTCAAC
1301 CTGCCAGTGC CCATGTTTGA GGACAACCTG GGGCCTCAGC TGTCAAAGC
1351 GGACAGGTGG CGGGAGTATG TCAGCCAGGT GTCTGGGGG AAGCTGAAGC
1401 GGAGGGTGAA GGGTTGGGCG CCGAGGGCGG GCCCCGGGGT GGGCGAGGCC
1451 CGGCTGGCCT CCACCGCAGT GGAGAGCGCA GGGGTATCAT CGGCGCCAGA
1501 GGGCACCAGC CCGGGGGATC GCTTGGGAAA CGCGGGAGAT GTTTGTGTGC
1551 CCCAGGCTTC CCCTAGGCGA TGGAGGCCCA AGATCAACTG GGCCTCCTTT
1601 CCGCGCCGCA GGAAGGAGCA GACAGCACC ACAGGTCAGG GGGCAGACAT
1651 CGAGGCTGAT CAGGGGGGAG AGGCTGCAGA TAGTCAAAG GAAGAGGCCA
1701 TAGCTGACCA GCGGGAAGGG GCTGCAGGTA ATCAGAGGGC TGGGGCCCCA
1751 GCTGACCAGG GGGCAGAGGC TGCAGATAAT CAGAGGGAAG AGGCTGCAGA
1801 TAATCAGAGG GCAGGGGCC CAGCTGAGGA GGGGGCAGAG GCTGCAGATA
1851 ACCAGAGGGA AGAGGCTGCA GATAATCAGA GGGCAGAGGC CCCAGCTGAC
1901 CAGAGGTCAC AGGGCACAGA TAACCACAGG GAAGAGGCTG CAGATAATCA
1951 GAGGGCGGAG GCCCAGCTG ACCAGGGGTC AGAGGTACA GATAATCAAA
2001 GGAAGAGGCG GTACATGAC CAGAGGGAAA GGGCCCCAGC TGTCCAGGGT
2051 GCAGATAATC AGAGGGCACA GGGCCGGGCT GGGCAGAGGG CAGAGGCTGC
2101 ACATAATCAG AGGGCAGGGG CCCCAGGTAT CCAGGAAGCT GAAGTCTCAG
2151 CTGCCCAAGG GACCACAGGA ACAGCTCCAG GAGCCAGGGC CCGGAAACAG
2201 GTCAAGACAG TGAGGTTCCA GACCCCTGGA CGCTTTTCTG GGTTTTGCAA
2251 GCGCCCGGAG GCCTTCTGGC AACTCCCCG GTTGCCAACC CTGCCCAAGA
2301 GAGTCCCCAG GGCAGGAGAG GTCAGGAACC TCAGGGTGCT GAGGGCCGAG
2351 GCCAGAGCAG AAGCTGAGCA GGGAGAGCAA GAAGACCAGC TGTGAGGTGA
2401 GGGCTAGAGA CAGCCCACGG GCCCTCCTC CAAGTGTGGG AGGGAGAGAT
2451 GCTCTGCCTC TGAACCTCAA AGTGGAGGTG GAGTGTCTGC CACGCTCCTA
2501 CCTAACCAAC CTCTTTATTC TCTTGTTAAA GTTTTGTTC TGCTTTGATT
2551 TTTTTTTAAA TTTTTTAGAG ACAGGGTCTC ACTCTGTGTC CCAGGCTGGA
2601 GTGCAGTGGC ATGATCATAA CCACTGCAG CCTCAAACCT CTGGCCTCAA
2651 GTGATCCTCC TGCTCGGCC TCCAAAATG CTGGGATTAC AGATGTGAGC
```

2701 CACCACACAC ACCATCTGAT TAAAAAAAAA AAATACTGAT TCCCTGTAGC
 2751 AACCCAAAAA AAAAAAAAAA AAAAA

BLAST Results

Entry HS164A7F from database EMBL:
 H.sapiens CpG island DNA genomic MseI fragment, clone 164a7, forward
 read cpg164a7.ft1a .
 Score = 740, P = 3.0e-25, identities = 150/151

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 779 bp to 2392 bp; peptide length: 538
 Category: similarity to known protein

1 MLQIGEDVDY LLIPREVRLA GGVWRVSKP ATKEAEFRER LTQFLEEEGR
 51 TLEDVARIME KSTPHPPQPP KKPKEPRVRR RVQQMVTPPP RLVVGTYDSS
 101 NASDSEFSDF ETSRDKSRQG PRGKKVRKM PVSYLGSKFL GSDLESEDDE
 151 ELVEAFLRRQ EKQPSAPPAR RRVNLPVPMF EDNLGPQLSK ADRWREYVSQ
 201 VSWGKLRVRV KGWAPRAGPG VGEARLASTA VESAGVSSAP EGTSPGDRLG
 251 NAGDVCVPQA SPRRWRPKIN WASFRRRRKE QTAPTGGQAD IEADQGGEEA
 301 DSQREEAIAD QREGAAGNQR AGAPADQGA EADNQREEEA DNQRAGAPAE
 351 EGAEAADNQR EEAADNQRAE APADQRSQGT DNHREEAADN QRAEAPADQG
 401 SEVTDNQREE AVHDQRE RAP AVQGADNQRA QARAGQRAEA AHNQRAGAPG
 451 IQEAEVSAAQ GTTGTAPGAR ARKQVKT VRF QTPGRFSWFC KRRRAFWHTP
 501 RLPTLPKRVP RAGEVRNLRV LRAEAREAE QGEQEDQL

BLASTP hits

Entry RNU67136_1 from database TREMBL:
 "A-kinase anchoring protein AKAP150"; Rattus norvegicus
 A-kinase anchoring protein AKAP150 mRNA, complete cds. Rattus
 norvegicus (Norway rat)
 Length = 714
 Score = 182 (64.1 bits), Expect = 1.2e-10, P = 1.2e-10
 Identities = 73/257 (28%), Positives = 104/257 (40%)

Alert BLASTP hits for DKFZphfbr2_22k3, frame 2

TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
 S-antigen gene, complete cds., N = 1, Score = 178, P = 3.7e-11

>TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
 S-antigen gene, complete cds.
 Length = 285

HSPs:

Score = 178 (26.7 bits), Expect = 3.7e-11, P = 3.7e-11
 Identities = 60/217 (27%), Positives = 97/217 (44%)

Query: 269 INWASFRRRKEQTAPTGGGA-DIEADQGGEEADSQRE-EAIADQ---REGAAGNQAGA 323
 +N + + + E G+G D E E +D+ E E I Q E A N+ AG+
 Sbjct: 47 LNKNGKGNKYEDLQEEGEGENDDEEHSNSESDNDEENEIIVGQDGSNEKAGSNEEAGS 106

Query: 324 PADQGAEEAADNQREEAADNQAGAPAEEGA--EAADNQ---EEAADNQRAEAPADQRS 377
 G+ E+A N++AG+ E G+ EA N+ EEA N++A + S
 Sbjct: 107 NEKAGSNEEAGSNEKAGSNEKAGSNEEAGSNEEAGSNEEAGSNEEAGSNEKAGSNEKAGS 166

Query: 378 QGTDNHREEAADNQRAEAPADQGEVTDNQREEAVHDQRE RAPAVQGADNQRAQAR--AG 435
 EEA N++A + + GS E+A +++ + G+ N++A + AG
 Sbjct: 167 NEKAGSNEEAGSNEKAGSNEEAGSNEKAGSNEKAGSNEEAGS-NEKAGSNEEAG 225

Query: 436 QRAEAAHNQAGA---PGIQEAEVSAAQGTGTGA-PGA 469

Report for DKFZphfbr2 22k3.2

```
[LENGTH]          538
[MW]               59402.19
[pI]               8.72
[HOMOL]            TREMBL:AF037364_1 gene: "MA1"; product: "paraneoplastic neuronal antigen MA1";
Homo sapiens paraneoplastic neuronal antigen MA1 (MA1) mRNA, complete cds. 4e-10
[PROSITE]          AMIDATION          1
[PROSITE]          MYRISTYL           12
[PROSITE]          CK2_PHOSPHO_SITE    11
[PROSITE]          PKC_PHOSPHO_SITE    6
[PROSITE]          ASN_GLYCOSYLATION   1
[KW]               All_Alpha
[KW]               LOW COMPLEXITY      18.03 %
```

SEQ MLQIGEDVDYLLIPREVL RAGGVVRVISKPATKEAEFRERLTQFL EEGRTLEDVARIME
SEG
PRD cccccccccccccccccccceeeeeecchhhhhhhhhhhhhccchhhhhhhhh

SEQ KSTPHPPQPPKKPEPRVRRRVQMVTPPRLVVGTYDSSNASDSEFSDFETSRDKSRQG
SEGxxxxxxxxxxxxxxxxxxxxx.....
PRD hccccccccccccccccchhhhhhhhhhhcccccceeeecccccccccccccccccccccc

SEQ PRRGKKVRKMPVSYLGSKFLGSDLESEDDEELVEAFRLRQEKQPSAPPARRRVNLPVPMF
SEGxxxxxxxxxxxxx.....
PRD cccccccccceeeccccccccccccchhhhhhhhhhhhhhhcccccchhhhhhhcccccc

```

SEQ      EDNLGPQLSKADRWREYVSQVSWGKLRVRVKGWAPRAGPGVGEARLASTAVESAGVSSAP
SEG      .....
PRD      cccccccchhhhhhhhhheeeccchhhhhhhccccccccchhhhhhhhhhhccccccc

SEQ      EGTSPGDRLGNAGDVCVPQASPRRWRPKINWASFRRRRKEQTAPTGGADIADQGGEAA
SEG      .....
PRD      cccccccccccccceeeccccccccccccchhhhhhhhhhhccccccchhhhhccccchh

SEQ      DSQREEAIADQREGAAGNQRAGAPADQGAEEADNQREEAADNQAGAPAEEGAEEAADNQ
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhh

SEQ      EEAADNQRAEAPADQRSQGTDNHREEAADNQRAEAPADQGEVTDNQREEAVHDQRERAP
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      AVQGADNQRAQARAGQRAEAAHNQRAGAPGIQAEVSAAGTTGTAPGARARKQVKTVERF
SEG      .....
PRD      hhccccchhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccchhhhhhhhhhh

SEQ      QTPGRFSWFCRRRAFWHTPRLPTLPKRVPRAGEVRNLRVLRAEAREAEQGEQEDQL
SEG      .....
PRD      cccccceehhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcc

```

Prosites for DKFZphfbr2_22k3.2

PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	273->276	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	499->502	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	112->116	CK2_PHOSPHO_SITE	PDOC00006
PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS00006	229->233	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	220->226	MYRISTYL	PDOC00008
PS00008	242->248	MYRISTYL	PDOC00008
PS00008	296->302	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	328->334	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	461->467	MYRISTYL	PDOC00008
PS00008	464->470	MYRISTYL	PDOC00008
PS00009	123->127	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_22k3.2)

DKFZphfbr2_22k8

group: brain derived

DKFZphfbr2_22k8 encodes a novel 172 amino acid protein without similarity to known proteins.

No informative BLAST results; no predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: /map="7"

Insert length: 2789 bp

Poly A stretch at pos. 2769, polyadenylation signal at pos. 2756

```
1 GGGGGAGCCA TGAGGCGCCA GCCTGCGAAG GTGGCGGCGC TGCTGCTCGG
51 GCTGCTCTTG GAGTGCACAG AAGCCAAAAA GCATTGCTGG TATTTCGAAG
101 GACTCTATCC AACCTATTAT ATATGCCGCT CCTACGAGGA CTGCTGTGGC
151 TCCAGGTGCT GTGTGCGGGC CCTCTCCATA CAGAGGCTGT GGTACTTCTG
201 GTTCTTCTGT ATGATGGGCG TGCTTTTCTG CTGCGGAGCC GGCTTCTTCA
251 TCCGGAGGCG CATGTACCCC CCGCCGCTGA TCGAGGAGCC AGCCTTCAAT
301 GTGTCTTACA CCAGGCAGCC CCCAAATCCC GGCCAGGAG CCCAGCAGCC
351 GGGGCGGCCC TATTACACTG ACCCAGGAGG ACCGGGGATG AACCTGTCTG
401 GGAATTCACG GGCAATGGCT TTCCAGGTCC CACCCAACTC ACCCCAGGGG
451 AGTGTGGCCT GCCCGCCCCC TCCAGCCTAC TGCAACACGC CTCGCGCCCC
501 GTACGAACAG GTAGTGAAGG CCAAGTAGTG GGGTGCCAC GTGCAAGAGG
551 AGAGACAGGA GAGGGCCTTT CCCTGGCCTT TCTGCTTCTG TTGATGTTCA
601 CTTCCAGGAA CGGTCTCGTG GGCTGCTAAG GGCAGTTCCT CTGATATCCT
651 CACAGCAAGC ACAGCTCTCT TTCAGGCTTT CCATGGAGTA CAATATATGA
701 ACTCACACTT TGTCTCTCTT GTTGCTTCTG TTTCTGACGC AGTCTGTGCT
751 CTCACATGGT AGTGTGGTGA CAGTCCCGGA GGGCTGACGT CTTACGGTGT
801 GCGTGACCAG ATCTACAGGA GAGAGACTGA GAGGAAGAAG GCAGTGTCTG
851 AGGTGCAGGT GGCATGTAGA GGGGCCAGGC CGAGCATCCC AGGCAAGCAT
901 CTTTCTGCCC GGGTATTAAT AGGAAGCCCC ATGCCGGGCG GCTCAGCCGA
951 TGAAGCAGCA GCCGACTGAG CTGAGCCGAG CAGGTTCATCT GCTCCAGCCT
1001 GTCTCTCTCGT CAGCCTTCCT CTTCCAGAAG CTGTTGGAGA GACATTACAG
1051 AGAGAGCAAG CCCCTTGTC TGTTCCTGTC TCTGTTTATA TCCTAAAGAT
1101 AGACTTCTCC TGCAACGCCA GGGAAAGGATA GCACGTGCAG CTCTACCCGC
1151 AGGATGGGGC CTAGAAATCAG GCTTGCCTTG GAGGCCTGAC AGTGATCTGA
1201 CATCCACTAA GCAAATTTAT TTAAATTCAT GGGAAATCAC TTCTGCCCC
1251 AAAGTGAAGC ATTGCATTTT GTGAGCTCTT GGTCTGATTT GGAGAAAGGA
1301 CTGTTACCCA TTTTCTTGGT GTGTTTATGG AAGTGCATGT AGAGCGTCTT
1351 GCCCTTTGAA ATCAGACTGG GTGTGTGTCT TCCCTGGACA TCACTGCCTC
1401 TCCAGGGCAT TCTCAGGCC GGGGGTCTCC TTCCCTCAGG CAGCTCCAGT
1451 GGTGGGTGCT GAAGGGTGCT TTCAAAACGG GGCACATCTG CCCGGGAAGT
1501 CACATGGACT CTTCCAGGGA GAGAGACCAG CTGAGGCGTC TCTCTCTGAG
1551 GTGTGTGTTG GTCTAAGCGG GTGTGTGCTG GGCTCCAAGG AGGAGGAGCT
1601 TGCTGGGAAA AGACAGGAGA AGTACTGACT CAACTGCACT GACCATGTTG
1651 TCATAATTAG AATAAAGAAG AAGTGGTCGG AAATGCACAT TCCTGGATAG
1701 GAATCACAGC TCACCCAGG ATCTCACAGG TAGTCTCCTG AGTAGTTGAC
1751 GGCTAGCGGG GAGCTAGTTC CGCCGCATAG TTATAGTGTG GATGTGTGAA
1801 CGCTGACCTG TCCTGTGTGC TAAGAGCTAT GCAGCTTAGC TGAGGCGCCT
1851 AGATTACTAG ATGTGCTGTA TCACGGGGAA TGAGGTGGGG GTGCTTATTT
1901 TTAATGAAC TAATCAGAGC CTCTTGAGAA ATTGTTACTC ATTGAACCTG
1951 AGCATCAAGA CATCTCATGG AAGTGGATAC GGAGTGATTT GGTGTCCATG
2001 CTTTTCACCT TGAGGACATT TAATCGGAGA ACCTCCTGGG GAATTTTGTG
2051 GGAGACACTT GGGAAACAAA CAGACACCTT GGGAAATGCA TTGCAAGCAC
2101 AGATGCTGCC ACCAGTGTCT CTGACCACCC TGGTGTGACT GCTGACTGCC
2151 AGCGTGGTAC CTCCATGCT GCAGGCCCTC ATCTAAATGA GACAACAAAG
2201 CACAAATGTT ACTGTTTACA ACCAAGACAA CTGCGTGGGT CCAAACACTC
2251 CTCTTCTCTC AGGTCATTTG TTTTGCATTT TTAATGTCTT TATTTTGTG
2301 AATGAAAAAG CACACTAAGC TGCCCTGGA ATCGGGTGCA GCTGAATAGG
2351 CACCCAAAAG TCCGTGACTA AATTCCGTTT GTCTTTTGA TAGCAAATTA
2401 TGTAAAGAGA CAGTGATGGC TAGGGCTCAA CAATTTTGTA TTCCCATGTT
2451 TGTGTGAGAC AGAGTTTGTG TTCCCTTGAA CTTGGTTAGA ATTGTGCTAC
2501 TGTGAACGCT GATCCTGCAT ATGGAAGTCC CACTTTGGTG ACATTTCTCTG
2551 GCCATTCTTG TTTCCATTGT GTGGATGGTG GGTGTGCCCC ACTTCTGGA
2601 GTGAGACAGC TCCTGGTGTG TAGAATTCCT GGAGCGTCCC TGGTTCAGAG
2651 TAAACTTGAA GCAGATCTGT GCATGCTTTT CCTCTGCAGC AATTGGCTCG
2701 TTTCTCTTTT TTGTTCTCTT TTGATAGGAT CCTGTTTCCT ATGTGTGCAA
```

BLAST Results

Medline entries

Peptide information for frame 1

```

1  MRRQPAKVAA LLLGLLECT EAKKHCWYFE GLYPTYIICR SYEDCCGSRG
51 CVRALSIQRL WYFFWLLMMG VLFCCGAGFF IRRRMYPPL IEEPAFNVSU
101 TRPPPNPGPG AQPQGPYYIT DPGGPGMNPV GNSTAMAFV PPNSPQGSVA
151 CQPPPPAYCNT PQQPYEOVVK AK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2 22k8, frame 1

PIR:S14970 extensin class I (clone w17-1) - tomato, N = 1, Score = 118,
P = 2.3e-07

>PIR:S14970 extensin class I (clone wl7-1) - tomato
Length = 132

HSPs:

Score = 118 (17.7 bits), Expect = 2.3e-07, P = 2.3e-07
Identities = 30/82 (36%), Positives = 35/82 (42%)

Query: 87 PPPLIEEPAFNVSYTRQPPNPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
PPP P Y + PP P P P P Y Y P P + P + P P SP
Sbjct: 32 PPPSPSPPP--PYYYSKSPSPSP--PPYYYSKSPPPDPSPPPYYYSKSPPPSPSP 87

```
Query:      147 GSVACPPPPPAYCNTPPPP--YEQV 168
              PPPP Y + PPPP  YE +
Sbjct:      88 PPSPSPPPPTYSSPPPPPPFYENI 111
```

Score = 104 (15.6 bits), Expect = 6.9e-06, P = 6.9e-06
Identities = 28/78 (35%), Positives = 34/78 (43%)

Query: 87 PPPLIEEPAFNVSYTRQPPNPGGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
 PP P P + Y K P P P P P P P Y Y P P P + P ++ P P
 Sbict: 1 PPSPSPPP---YKSPSPSPSP---PPYYKSPSPSPSP---PPYYKSPSP-PPS 51

```
Query: 147 GSVACPPPPAYCNTPPPP 164
      S P P P P Y + P P P P
Sbjct: 52 PS---PPPPYYYKSPPPP 66
```

Score = 102 (15.3 bits), Expect = 1.1e-05, P = 1.1e-05
Identities = 30/78 (38%), Positives = 33/78 (42%)

Query: 87 PPPLIEEPAFNVSYTRQPPNPGGAQQGPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
 PPP P Y + PP P P P P Y Y P P + P S + PP P
 Sbjct: 48 PPPSPSPPP---PYYYSKPPPPDPSP---PPYYYSKPPPPSPSPPPSPS-----PP-PPT 97

```

Query:      147 GSVACPPPPAYCNTPTPP 164
           S   PPPP Y N P PP
Sbjct:      98 YSSPPPPPPPFYENIPLPP 115

  Score = 95 (14.3 bits), Expect = 2.4e-04, P = 2.4e-04
  Identities = 24/61 (39%), Positives = 29/61 (47%)

Query:      104 PPNPGPGAQQGPPYYTDPGGPGMNVGNSTAMAFQVPPNSPQGSVACPPPPAYCNTPTPP 163
           PP+P P      P P YY P P +P      ++ PP P S   PPPP Y +PPP
Sbjct:      1  PPSPSP----PPPYYYKSPPPPSPSP---PPPYYYKSPP-PPSPS---PPPPYYKSPPP 49

Query:      164 P 164
           P
Sbjct:      50 P 50

  Score = 68 (10.2 bits), Expect = 4.2e+00, P = 9.8e-01
  Identities = 24/69 (34%), Positives = 29/69 (42%)

Query:      87 PPPLIEEPAFNVSYTRQPP---NPGPGAQQGPPYYTDPGGPGMNVGNSTAMAFQVPPN 143
           PPP   P      Y   PP +P P + P PP Y+ P P   P + +   PP
Sbjct:      63 PFPDPSPPPPYYYKSPPPPSPSPPPSPSPPPPTYSSPPPPP--PFYENIPL----PPV 116

Query:      144 SPQGSVACPPPP 155
           S A PPPP
Sbjct:      117 IGV-SYASPPPP 127

```

Peptide information for frame 3

ORF from 0 bp to 368 bp; peptide length: 123
Category: questionable ORF
Classification: unset

```

1 GSHEAPACEG  GGAAARAALG  VHRSQKALLV  FRRTLSNLLY  MPLLRGLLWL
51 QVLCAGPLHT  EAVVLLVPSD  DGRAFLRLSR  LLHPEAHVPP  AADRGASLQC
101 VLHOAAPKSR  PRSPAAGAAL  LH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2 22k8, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 22k8, frame 1

Report for DKFZphfbr2 22k8.1

```

[LENGTH]                172
[MW]                     19194.47
[pI]                     8.77
[KW]                     SIGNAL_PEPTIDE 23
[KW]                     TRANSMEMBRANE 1
[KW]                     LOW_COMPLEXITY 27.33 %

SEQ      MRRQPAKVAALLGLLLECTEAKHKCWYFEGLYPTYIICRSYEDCCGSRCVRLSIQRL
SEG      .....XXXXXXXXX
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhhhccccccccccceeeccccccccccchhhhhhhhhh
MEM      .....

SEQ      WYFWFLLMMGVLFCCGAGFFIRRRMYPPLIEEPAFNVSYTRQPPNPFGPAQQPGPPYYT
SEG      .....XXXXXXXXXXXXXXXXXXXX
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccceeecccccccccccccccccc
MEM      ....MMMMMMMMMMMMMMMMMM

SEQ      DPGGPGMNPVGNSTAMAFQVPPNSPQGSVACPPPPAYCNTPPPPYEQVVKAK
SEG      XXXXXXXX.....XXXXXXXXXXXXXXXXXXXX
PRD      ccccccccccccccccccecccccccccccccccccccccccccccccccccccccc
MEM      .....

```

(No Prosite data available for DKFZphfbr2 22k8.1)

(No Pfam data available for DKFZphfbr2_22k8.1)

Pedant information for DKFZphfbr2_22k8, frame 3

Report for DKFZphfbr2_22k8.3

[LENGTH] 122
[MW] 12854.08
[pI] 10.27
[KW] All_Alpha
[KW] LOW_COMPLEXITY 25.41 %

SEQ GSHEAPACEGGGAAARAALGVHRSQKALLVFRRTLSNLLYMPLLRGLLWLQVLCAGPLHT
SEGXXX
PRD cccccccccchhhhhhhccccchhhhhhhhhhhhhhhccccccccchhhhhhhcccccc

SEQ EAVVLLVPSDDGRAFLRSRLHPEAHVPPAADRGASLQCVLHQAAPKSRPRSPAAGAAL
SEGXXX
PRD cceeeeeccccchhhhhhhccccccccccccccccchhhhhhhccccccccchhhhhhc

SEQ LH
SEG ..
PRD cc

(No Prosite data available for DKFZphfbr2_22k8.3)

(No Pfam data available for DKFZphfbr2_22k8.3)

DKFZphfbr2_23b10

group: nucleic acid managment

DKFZphfbr2_2b10 encodes a novel 580 amino acid protein with strong similarity to rat RNA helicase HEL117.

HEL117 is a DEAD/H box helicase, which co-localises with a splicing factor and thus seems to be involved in splicing.

The new protein can find application in modulation of splicing.

strong similarity to rat RNA helicase HEL117

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2905 bp

Poly A stretch at pos. 2885, no polyadenylation signal found

```

1  GGGGGCTCCG CTCCGCACCA CCAACCCCGG GCCCGAGTCC TGACGAGCGG
51  GTCAGGGCTT GTCGGGCGGA AGCCTGGCCT GGAGCCTGGA AGGGGGAGAC
101 GGCCCGAGCG GGAGCGGGAG CGGACGCGGC CTCAGTCCTG CGCGGAATAT
151 TGAAGGATGT TTGTTCCAAG ATCTCTAAAA ATCAAGAGGA ATGCTAATGA
201 TGATGGCAAA AGTTGTGTGG CTAAGATAAT TAAACCAGAC CCAGAAGACC
251 TTCAGTTTGA CAAAAGCAGA GATGTTCCCG TTGATGCTGT AGCTACAGAA
301 GCAGCCACAA TAGACAGGCA CATCAGCGAA TCATGCCCTT TCCCCAGCCC
351 AGGTGGCCAG TTGGCAGAGG TTCATTCAGT AAGTCCCGAG CAGGGTGCGA
401 AGGACAGCCA TCCTTCTGAA GAGCCCGTTA AGTCATTTTC CAAAACACAG
451 CGCTGGGCAG AACCAGGGGA ACCCATCTGT GTTGTCTGTG GTCGTTATGG
501 AGAGTATATC TGTGATAAGA CAGATGAAGA TGTGTGTAGT TTGGAGTGTA
551 AAGCGAAACA TCTTCTACAA GTTAAGGAAA AGGAAGAGAA ATCAAAACTC
601 AGCAATCCAC AGAAGGCTGA TTCTGAGCCA GAGTCTCCAC TGAATGCTTC
651 CTATGTCTAC AAAGAGCACC CCTTTATTTT GAACCTTCAG GAAGACCAGA
701 TTGAAAATCT TAAACAGCAG CTGGGAATTT TAGTTCAAGG GCAAGAAGTC
751 ACCAGGCCCA TTATTGACTT TGAACATTGT AGTCTCCCTG AGGTCTTAAA
801 TCACAACCTT AAGAAATCAG GCTATGAGGT GCCAACTCCC ATTCAAATGC
851 AGATGATTCC TGTGGGACTT CTGGGAAGAG ACATTCTGGC CAGTGCAGAT
901 ACTGGCTCAG GAAAAACAGC TGCTTTTCTT CTTCCTGTTA TCATGCGAGC
951 TTTATTTCGAG AGCAAAACTC CATCTGCGCT CATTCTTACA CCAACCAGAG
1001 AGTTAGCCAT TCAGATAGAG AGACAAGCTA AAGAATTGAT GAGTGGCCTG
1051 CCACGCATGA AAACGTGTCT TCTTGTAGGG GGCTTACCCT TACCCCCACA
1101 GCTTTATCGT CTGCAACAAC ATGTTAAGGT TATCATAGCA ACCCCTGGGC
1151 GACTTCTGGA TATAATAAAG CAGAGCTCTG TAGAACTCTG TGGTGTAAGG
1201 ATTGTGGTAG TAGATGAAGC TGATACCATG TTAAGATGGG GTTTTCAACA
1251 ACAAGTGCTT GACATTTTGG AAAACATTCC TAATGATTGT CAGACCATTT
1301 TGGTTTCAGC CACAATTCCA ACTAGCATAG AACAGTAGC AAGCCAGCTT
1351 CTGCATAATC CTGTGAGAAT TATCACTGGA GAAAAGAACC TACCTTGTGC
1401 CAATGTACCT CAGATTATTT TGTGGGTAGA AGACCCAGCC AAAAAAGAAA
1451 AATTATTTGA AATTTTAAAT GATAAGAAAC TCTTTAAGCC TCCAGTGTTA
1501 GTATTGTGGG ACTGCAAACT AGGAGCAGAT CTTTGAAGT AAGCCGTTCA
1551 GAAAATCACA GGGCTGAAAA GCATATCTAT ACATTGGGAG AAGTCGCAAA
1601 TAGAAAGGAA AAACATATTG AAGGGATTAC TTGAAGGAGA CTATGAAGTT
1651 GTAGTGAGCA CAGGAGTCTT GGGACGAGGC CTAGACTTGA TCAGTGTCAG
1701 GCTGGTTGTC AATTTTGATA TGCCTTCAAG TATGGATGAG TATGTCCATC
1751 AGGAAAATAC CTACAAGTCT ACTTGGAGGA ATCCCAGCA TTTTCAACAG
1801 GATGTCAGAA TGACCTTGGG CTATGTTGGC AAAGCACAAAT GGAAGAAGA
1851 CAACCAATTG AAGGTCAAAC TAGGCCTTAA AAAAAATTGT TCTTCTTAAA
1901 TGAAACTTTA TGTAAGACCC AAGCTTCCTT TATGTAAAAA TAGGATACTC
1951 ACTAGGCTTT GGGGCTGACA ATGGTTTTTA AATCTTGCTA ATCTTCCCTG
2001 GAATGAAACC AGCATGACTC AAAGAGAAAA AGAGAGTCTA TAATATTTTC
2051 TAATCCCTGA GTTCTTTTCT TTATATATTA AAAAGGATTA TTAGGCTGGG
2101 TGTGGTGGCT CACGCCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGGGAGT
2151 GGATCACCTG AGTTCGAGAC CAGCCTAAC AACATGGAGA AACCTGTCT
2201 CTAATAAAAA TACAAAATTA GCCAGGCGTG GTGGCGCATG CCTGTAATCC
2251 CAGCTACTCA GGAGGCTACA GCAGGAGAAT TGCTTGAAC TCGGAGGCAG
2301 AGCCAAGATC GCACCACTGC ACTCCAGCCT GGGCAACAAG AGTGAAACTC
2351 TGCTCTAAAA TAATATTAAT GATAATAATA ATAATAATAA TAGGGATTAC
2401 TTGCATAATT GTTCTTTTAA AATTATTGGC AGTATTGCTG AATGTATTTA
2451 GATTTTTTCA CCAAGTGACA ACAACTGAAT TCATAAAGAT TCATCAACAA
2501 GACCTGATAA AAAAAAATGT AAGCATATTA TAGTGATAC TTCCAAGACT
2551 CTTGGTCTAA CATGTATTAG AAAGCAGAAG GAGCCAGGCG ACAGGGGCTC
2601 CCGCCGGTAA TCCCAAAGCT TTGGGAAGCC AAGGCAGGTG GATCGCTTGA
2651 GCTCAGGAGT TAGAGACCAG CCTGGGCAAC ATGGTGAAAT CCCGTACCCA

```

```

2701 CAAAAAATG CAAAAATTAA CTGGGCGTGG TGGCATGCAC CTGTAGTCCC
2751 AGCTACTCTG GAGGCTGAGG TGAGGGGAAT CACCTGAGCC GGGGGAATCA
2801 CCTGAGCCCA GGAAGTTGA GGCTGCTGTG AGCCATGGTC ATGACACTGC
2851 CCTCCAGCCT GGACAACAGA TTGAGACCCT GTCTCAAAAA AAAAAAAAAA
2901 AAAAA

```

BLAST Results

No BLAST result

Medline entries

Medline:

A putative mammalian RNA helicase with an arginine-serine-rich domain

Peptide information for frame 1

ORF from 157 bp to 1896 bp; peptide length: 580
 Category: strong similarity to known protein
 Prosite motifs: ATP_GTP_A (247-255)
 LEUCINE_ZIPPER (298-320)

```

1 MFVPRSLKIK RNANDDGKSC VAKIIKPDPE DLQDKSRDV PVDVATEAA
51 TIDRHISESC PFPSPGGQLA EVHSVSPEQG AKDSHPSEEP VKSFSKTQRW
101 AEPGEPICVV CGRYGEYICD KTDEDVCSLE CKAKHLLQVK EKEEKSKLSN
151 PQKADSEPEL PLNASYVYKE HPFILNLQED QIENLKQQLG ILVQGQEVTR
201 PIIDFEHCSL PEVLNHNLLK SGYEVPTPIQ MQMIPVGLLG RDILASADTG
251 SGKTA AFLLP VIMRALFESK TPSALILTPT RELAIQIERQ AKELMSGGLPR
301 MKTVLLVGG LPLPQLYRLQ QHVKVIIATP GRLLDIKQS SVELCGVKIV
351 VVDEADTMLK MGFQQQVLDI LENIPNDCQT ILVSATIPTS IEQLASQLLH
401 NPVRIITGEK NLPKANVRQI ILWVEDPAKK KKLFEILNDK KLFKPPVLVF
451 VDCKLGLADLL SEAVQKITGL KSIHSEKS QIERKNILKG LLEGDYEVVV
501 STGVLGRLGD LISVRLVVF DMPSSMDEYV HQENTYKSTW RNPQHQQQDV
551 RMTLGYVGKA QWEDNQLKV KLGLKNCSS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_23b10, frame 1

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 615, P = 1.6e-60

TREMBL:AB018344_1 gene: "KIAA0801"; product: "KIAA0801 protein"; Homo sapiens mRNA for KIAA0801 protein, complete cds., N = 1, Score = 615, P = 2.8e-59

TREMBL:CEF01F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1., N = 2, Score = 365, P = 1.9e-58

TREMBL:AF083255_1 product: "RNA helicase-related protein"; Homo sapiens RNA helicase-related protein mRNA, complete cds., N = 2, Score = 556, P = 1.5e-57

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces pombe), N = 1, Score = 591, P = 1.6e-57

>PIR:A57514 RNA helicase HEL117 - rat
 Length = 1,032

HSPs:

Score = 615 (92.3 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
 Identities = 140/394 (35%), Positives = 236/394 (59%)

```

Query: 144 ESKLSNPQKADSEPEPLNASYVYKEHPFILNLQEDQIENLKQQL-GILVQGQEVTRPI 202
      ++ KL P P ++ Y E P + +++++ + ++ GI V+G+ +PI
Sbjct: 313 QQRKLLPEVDHGKIEYEPFRKNF-YVEVPELAKMSQEEVNVFRLEMEGITVKGKGC PKPI 371

Query: 203 IDFEHCSLPEVLNHNLLKSGYEVPTPIQMIPVGLLGRDILASADTGSGKTA AFLLPV- 261

```

```

      + C + + ++LKK GYE PTPIQ Q IP + GRD++ A TGSGKT AFLLP+
Sbjct: 372 KSWVQCISGMKILNSLKKHGYEKPITQIQAIPAINSGRDLIGIAKTGSGKTIAFLLPMF 431
Query: 262 --IM--RALFESKTPSALILTPTRELAIQIERQAKELMSGGLPRMKTIVLLVGGGLPLPPQLY 317
      IM R+L E + P A+I+TPTRELA+QI ++ K+ L ++ V + GG + Q+
Sbjct: 432 RHIMDQRSLEEGEGPIAVIMTPTRELAQITKECKKFSKTLG-LRVVCVYGGTGISEQIA 490
Query: 318 RLQQHVKVIIATPGRLLDIIKQSS---VELCGVKIVVVDEADTMLKMGFQQQVLDILENI 374
      L++ ++I+ TPGR++D++ +S L V VV+DEAD M MGF+ QV+ I++N+
Sbjct: 491 ELKRGAEIIVCTPGRMIDMLAANSGRVTNLRRVTVVVLDEADRMFDMGFEPQVMRIVDNV 550
Query: 375 PNDCQTILVSATIPSTIEQLASQLLHNPVRIITGEKNLPCANVRQIILWVEDPAKKKKLF 434
      D QT++ SAT P ++E LA ++L P+ + G +++ C++V Q ++ +E+ K KL
Sbjct: 551 RPDRTQVMFSATFPRAEALARRILSKPIEVQVGGRSVVCSDVEQQVIVIEEEKKFLKLL 610
Query: 435 EILNDKKLFKPPVLVFDCKLGDLLSEAVQKITGLKSISIHSEKSQIERKNILKGLLEG 494
      E+L + V++FVD + AD L + + + + +S+H Q +R +I+ G
Sbjct: 611 ELLGHVQE-SGSVII FVDKQEHADGLLDLMRAS-YPCMSLHGGIDQYDRDSIINDFKNG 668
Query: 495 DYEUVVSTGVLGRGLDLISVRLVVFNDMPSSMDEYVHQ 532
      +++V+T V RGLD+ + LVVN+ P+ ++YVH+
Sbjct: 669 TCKLLVATSVAAAGLDVKHLILVVNYSCPNIHEDYVHR 706

Score = 37 (5.6 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
Identities = 13/36 (36%), Positives = 17/36 (47%)

Query: 132 KAKHLLQVKEKEE---KSKLSNPQKADSEPEPLNA 164
      KA++ + KEK E SK K D E E +A
Sbjct: 113 KAENRSRSKEKAEGGDSKSKKKDKDDKEDEKEKDA 148

```

Pedant information for DKFZphfbr2_23b10, frame 1

Report for DKFZphfbr2_23b10.1

```

[LENGTH] 580
[MW] 64572.24
[pI] 6.13
[HOMOL] TREMBL:CEF01F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1. 8e-61

[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-53
[FUNCAT] 04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-53
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 5e-53
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-49
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YOR204w] 2e-49
[FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 2e-46
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 3e-43
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 4e-39
[FUNCAT] l genome replication, transcription, recombination and repair [H.
influenzae, HI0892] 3e-35
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 6e-34
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YOR046c] 3e-32
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 8e-30
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 5e-23
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 1e-16
[FUNCAT] r general function prediction [M. jannaschii, MJ1401] 5e-11
[FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 1e-06
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YMR190c] 1e-06
[BLOCKS] BL001158 Eukaryotic RNA polymerase II heptapeptide repeat proteins
[BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW] nucleus 6e-53
[PIRKW] RNA binding 9e-52
[PIRKW] DEAD box 2e-43
[PIRKW] transmembrane protein 1e-21
[PIRKW] DNA binding 5e-48
[PIRKW] ATP 4e-57
[PIRKW] purine nucleotide binding 2e-43
[PIRKW] P-loop 4e-57
[PIRKW] hydrolase 6e-42
[PIRKW] protein biosynthesis 2e-43
[PIRKW] ATP binding 2e-50
[SUPFAM] WW repeat homology 1e-49
[SUPFAM] translation initiation factor eIF-4A 2e-43
[SUPFAM] DEAD/H box helicase homology 4e-57
[SUPFAM] recQ helicase homology 8e-06

```


PS00008	195->201	MYRISTYL	PDOC00008
PS00008	250->256	MYRISTYL	PDOC00008
PS00008	490->496	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008
PS00017	247->255	ATP_GTP_A	PDOC00017
PS00029	298->320	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphfbr2_23b10.1

HMM_NAME DEAD and DEAH box helicases

HMM	*gLpPWILRnIyeMGFEkPTPIQQqAIPiIleGRDVMACAQTGSGKTAAAF +LP+ + N+++ G+E PTPIQ+Q IP+ L GRD++A A TGSGKTAAAF	
Query	209 SLPEVLNHNLLKSGYEVPTPIQMIPVGLLGRDILASADTGSGKTAAAF	257
HMM	lIPMLQHIIdwdPWpqpPQdPrALILAPTRELAMQIQEEcrkFgkHMngIR L+P++ + + + ++P ALIL+PTRELA+QI++++++ + ++ ++	
Query	258 LLPVIMRALFES--KTPS---ALILPTRELAIQIERQAKELMSGLPK	302
HMM	ImcIYGGtnMRdQMRmLeRGpPHIVATPGRLIDHIERgtldLDrIeMLV ++++GG+++ +Q+ +L++ + ++IATPGRL+D+I++ ++ L ++++V	
Query	303 TVLLVGGLPLPPQLYRLQQHV-KVIIATPGRLLDIIKQSSVELCGVKIVV	351
HMM	MDEADRMLDMGFIDQIRrIMrqIPMpwnRQTMFSATMPdeIqELARrFM DEAD ML MGF++Q+ +I+ IP + QT++ SAT+P +I++LA ++	
Query	352 VDEADTMLKMGFQQQVLDILENIP--NDCQTILVSATIPTSIEQLASQLL	399
HMM	RNPIRInIdMdElTtnEnIkQwYiyVerEMWKfdeLcrLIe* +NP+RI+ ++++L N++Q++ +VE + K +L+++++	
Query	400 HNPVRIITGEKNLPCA-NVRQIILWVE-DPAKKKKLFEILN	438

HMM_NAME Helicases conserved C-terminal domain

HMM	*EileeWLknl.GIrvMYIHGdMpQeERdeIMddFNnGEynVLicTDVgg ++L+E ++ G++ ++IH+ ++Q ER +I++ +G+Y V ++T V+G	
Query	458 DLLSEAVQKITGLKSISIHSEKSQIERKNILKGLLEGDYEVVVSTGVLG	506
HMM	RGIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG* RG+D+++V++V+N+DMP +++ Y++ + T +	
Query	507 RGLDLISVRLVNVFDMPSMDEYVH-QENTYKST	539

DKF2phfbr2_23b21

group: signal transduction

DKF2phfbr2_23b21.1 encodes a novel 193 amino acid protein which is nearly identical to bovine neurocalcin.

Neurocalcin is a Ca(2+)-binding protein with three putative Ca(2+)-binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca²⁺ dependent activation of guanylate cyclase.

The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

nearly identical to bovine neurocalcin

complete cds complete cDNA
EST hits

Sequenced by AGOWA

Locus: /map="574.6 cR from top of Chr8 linkage group"

Insert length: 3300 bp

Poly A stretch at pos. 3279, polyadenylation signal at pos. 3249

```

1 GGGGAGAATC TGGTGGATGC TGGACCTTGC TGCTGCTGCT ACTGCTGTTT
51 CCAGGGGCTG CAGAGCATGG ACTGTAAAT CTGCACTTC TTCTGAGTGA
101 GCTGAATTCT TGCCGCCAGG ATGGGGAAAC AGAACAGCAA GCTGCGCCCG
151 GAGGTCATGC AGGACTTGCT GGAAAGCACA GACTTTACAG AGCATGAGAT
201 CCAGGAATGG TATAAAGGCT TCTTGAGAGA CTGCCCCAGT GGACATTGT
251 CAATGGAAGA GTTTAAGAAA ATATATGGGA ACTTTTCCCT TTATGGGGAT
301 GCTTCCAAAT TTGCAGAGCA TGTCTCCGCG ACCTTCGATG CAAATGGAGA
351 TGGGACAATA GACTTTAGAG AATTCATCAT CGCCTTGAGT GTAACCTCGA
401 GGGGGGAAGCT GGAGCAGAAG CTGAAATGGG CCTTCAGCAT GTACGACCTG
451 GACGGAAATG GCTATATCAG CAAGGCAGAG ATGCTAGTGA TCGTGCAGGC
501 AATCTATAAG ATGGTTTCCT CTGTAATGAA AATGCCTGAA GATGAGTCAA
551 CCCCAGAGAA AAGAACAGAA AAGATCTTCC GCCAGATGGA CACCAATAGA
601 GACGGAAAAC TCTCCCTGGA AGAGTTCATC CGAGGAGCCA AAAGCGACCC
651 GTCCATTGTC CGCCTCCTGC AGTGCAGACC GAGCAGTGCC GGCCAGTTCT
701 GAGCCCTGCG CCCACCAATC GAATTGTAGA GCTGCTTGTG TTCCCTTTTG
751 ATTCTTCTTT TTAACAATTT TTTTTTTTTT TTGCCAAACA ATATCAATGG
801 TGATGCCGTC CCCTGTGCGG TCTGATGCGC CTTCCTCCGT GACGCCCTCA
851 GCCTCTTTTG TCGTGGATGC TTCGTGGGAA TGCCAGAGC CCCAGTGTGC
901 TTGTGGAGAG CATGGACAGA CTCGTGGTGT TTCATTGTTT GATGATTTTT
951 AATCGTTACT ATTATTCTTT TTTATTCTAA TGTCTCTGTT CTAAAACGTA
1001 AGACTCGGGG GTTGGGGCAA AAGAAGGGAA ACCCATCCAG TCCTGTGATT
1051 CTATTGCAAG CTTCAAGGGG CTTTTGTTTG AAAGACAAAA CTCCCCACCT
1101 GGGTCTGTTG TCACACGTGC CGTAGGGGTG ATGGATGGCA CCGGATGCTG
1151 GATTCCCAA GAACAAGTTA CCCTCTGGGG TGAGGCTATT CCAGCGAGCT
1201 GGGACATTTT CCCATGGGGG CCCACTCCCC TCTCTTCCCC AGCAGGCTGT
1251 AGTTTCTAAG CTGTGAACAT TTCAAGATAA ATTAACAGAG GAGAGGAAAA
1301 AGATGGCTCA GCTATTTTTT CACAGGTTTA CACTAGTTGA GCTAATATGC
1351 GTGCTTTTGG AAATTAACA CAAATGGTAA CATATTCCAA AACCAGACCC
1401 ATCTTTGTGC CTATTGTGAT AAAATAAAAA GACGGCTGTA TATAACATAT
1451 TGGGTAAATG AGACCAAATT AAGTGTTTTG CCTTGTTTAA ATGAAATGCA
1501 TGTTTAGTGA GCACTAATAC AATCTTATTC CAGAAGACTG TTTTATAGTAG
1551 CTTATTGTGA AGTAAGACAA CTATAATGAA TGTCTGTCTT GTTTGGAAGT
1601 CATATCTGTC TTGACACAAA TGTACCAATC GACAAGTATA TTTTATATAT
1651 TCCATAAAAA TACAAAGTAA CCCTGACTAG GGCCCACTT TAATTTTGAA
1701 TGCATTTCCA GAGTGGCCAT GCCTAGAGGG CAGATGCAGA GCAGGTGGTA
1751 GTGGGACAGG ACAATTGGAG CACAGGAATG TTAACATGTA TGACAGGGGA
1801 CCAGTAGGGT GGTTCCTCTC TCAGGCCAG CAGCCCATG ACAGCATTAG
1851 ACTGGCGGCA TGGTGCTTTT CTGAGCAGAT CAATACTCTG CAGACTCGAA
1901 AAAACATCAC ATACATTCTT GGAACCTCCC AGTGGTTTAA TCTATGTGCA
1951 TGGTTAGGGA GCCAGGCCCTG GAATATTCAG TTCCCTGCC CCTGTTAAAG
2001 AATCAGAGGT TGGGCAGTCA TCAAATTCAT CATAAAGACA TGGGCAAGTG
2051 TGTCTGTGGT TTCCAAGGCC CCCCTATGGA GAATCCAAAA GTATTTTCCA
2101 TTGCCGTGCT CTTTGAATGC AGACTTCTAT TTCCAGAAGT GACAGACCAA
2151 GTCTGAGTTG CTGTTTGGTC TGGTGACCTC AGACACACTA ATTTGAATTG
2201 AAAGCTAAGA GTAAAAATTT GCTGGTTACA GGCGAGTCAT ACTCTTGCAA
2251 CTAGTTAGCA AAGGGAGGCC CAAATTCCTA AGGTTGTGTA TGGGGAACCT
2301 GCCACTAAGA GAAGGCAGAG AGGTCCCTAG TGGGTATATT TGCTGCCAAG
2351 CCACTTGCCA AAGAAGAGGA ACCACAGAAA GAGAGACATC ATGACCAGGA
2401 GAAAAATGTG ACTAGACATG CTAACCTCCA GGTTTTATA TATGACTTGA
2451 GTCTGCTGTA ATTGGCAGCA GAAATCCAAA TTTGTATGGT AGACCAAAAA
2501 GAACCAATC CATAGGGTGA AATTTTGAGA CTAGACTCT GTAAAAATAA

```

```

2551 TCCTAGTCTT CCTCCAGGGG TCAGTTCCTC ACAGTGGTTC TGTACCAAAA
2601 CTTGCCAAAT TCCTCCATGG CCAAGTGTTA AAATCTGTGT TTGGAAAATA
2651 GCGAATTAAC CTAAGACACA GAAGGCAGAC TGGGTGAGGA GACCTAGCAT
2701 GCCCTATTGG CAGTGCTCAG GAGCTGCATC CCACTTTTCC CTGCTCTGAA
2751 TCGAAGTCCT AGTTCCTTCC TTTGATTCTC CTTTGGTAGG TGGAATCAGT
2801 TAATGTTTTG AGAAACCTGC CTGGGCTCTG CCCTTAGTCA TGACATCTCG
2851 CTGAGCCAGA CCCACTCTGT TCCTTGAAC CTAGAGCTGG AGTGAGGAGT
2901 AGAGGTCTCC GGCTATTCCA GAAAGAAAAG TGAGCCACAT GCAGGCTGAT
2951 GAATGCCGAC ACTTCCAGAA TGTATAGAAA TAGTCCCTGT CCTGGCCTGC
3001 CACTGACCCT GTCTGTATT TCTCGGAGGT TGTTTTTCTC CTTCTCCTTC
3051 CCAGGAAGGT CTTTGTATGT CGAATCCAGT GCACTCAAGT TTGGCCAAGG
3101 GACTCCACAG CACCCAGAGG ACTGCATGCC TCAAGGTTA TGTCACTCCT
3151 CTGCTGGGCT GTTCATTGTC ATTGCTGTGT TCAGGGACCT TTGGAAATAA
3201 AACCTGTTCT GTCCCAAATA AAACCAGCCT GTGATGTCA AGGGACTGGA
3251 ATAAAGTGGC TTACGACCTG AAGGATTCTA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HS431350 from database EMBL:
human STS WI-15914.
Score = 1308, P = 3.1e-53, identities = 276/285

Entry HSG19929 from database EMBL:
human STS A002C26.
Score = 926, P = 1.5e-35, identities = 186/187

Entry AF052142 from database EMBL:
Homo sapiens clone 24665 mRNA sequence.
Score = 7378, P = 0.0e+00, identities = 1482/1487
3' UTR

Medline entries

93247712:
Neurocalcin family: a novel calcium-binding protein abundant in bovine central nervous system.

94045365:
Distinct regional localization of neurocalcin, a Ca(2+)-binding protein, in the bovine adrenal gland.

96407688:
Crystallization and preliminary X-ray crystallographic studies of recombinant bovine neurocalcin delta.

96066284:
Distribution pattern of three neural calcium-binding proteins (NCS-1, VILIP and recoverin) in chicken, bovine and rat retina.

Peptide information for frame 1

ORF from 121 bp to 699 bp; peptide length: 193
Category: strong similarity to known protein
Prosite motifs: EF_HAND (73-86)
EF_HAND (109-122)
EF_HAND (157-170)

```

1 MGKQNSKLRP EVMQDLLEST DFTEHEIQEW YKGFLRDCPS GHLSMEEFKK
51 IYGNFFPYGD ASKFAEHVFR TFDANGDGTI DFREFIIALS VTSRGKLEQK
101 LKWAFSMYDL DNGYISKAE MLVIVQAIYK MVSSVMKMPD DESTPEK RTE
151 KIFRQMDTNR DGKLSLEEFI RGAKSDPSIV RLLQCDPSSA GQF

```

BLASTP hits

Entry JH0616 from database PIR:
neurocalcin (clone pCalN) - bovine

Score = 1001, P = 5.2e-101, identities = 192/193, positives = 192/193

Entry GGU91630_1 from database TREMBL:

product: "neurocalcin"; Gallus gallus neurocalcin mRNA, complete cds.

Score = 998, P = 1.1e-100, identities = 191/193, positives = 192/193

Entry NECD_BOVIN from database SWISSPROT:

NEUROCALCIN DELTA.

Score = 996, P = 1.8e-100, identities = 191/192, positives = 191/192

Entry S47565 from database PIR:

BDR-1 protein - human

Score = 934, P = 6.6e-94, identities = 174/193, positives = 187/193

Entry I50676 from database PIR:

gene Rem-1 protein - chicken >TREMBL:GGREM1_1 gene: "Rem-1"; G.gallus

rem-1 mRNA

Score = 933, P = 8.4e-94, identities = 174/193, positives = 186/193

Alert BLASTP hits for DKFZphfbr2_23b21, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23b21, frame 1

Report for DKFZphfbr2_23b21.1

```
[LENGTH]      193
[MW]           22215.30
[pI]           5.35
[HOMOL]        PIR:JH0616 neurocalcin (clone pCalN) - bovine 1e-109
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YDR373w] 3e-54
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
                [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       03.01 cell growth [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       13.04 homeostasis of other ions [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       08.19 cellular import [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR109c]
0.001
[FUNCAT]       10.02.99 other morphogenetic activities [S. cerevisiae, YBR109c] 0.001
[FUNCAT]       30.05 organization of centrosome [S. cerevisiae, YBR109c] 0.001
[BLOCKS]       BL00018
[SCOP]         dlrec_ 1.34.1.5.18 Recoverin [bovine (Bos taurus) 8e-55
[SCOP]         dljsa_ 1.34.1.5.17 Recoverin [human (Homo sapiens) 5e-58
[SCOP]         dltcob_ 1.34.1.5.16 Calcineurin regulatory subunit (B-chain 1e-06
[SCOP]         d2mysc_ 1.34.1.5.15 Myosin Regulatory Chain [chicken (Gallu 2e-29
[SCOP]         dlscmc_ 1.34.1.5.14 Myosin Regulatory Chain [bay scallo 5e-33
[SCOP]         d2mysb_ 1.34.1.5.13 Myosin Essential Chain [chicken (Gallu 4e-26
[SCOP]         dlscmb_ 1.34.1.5.12 Myosin Essential Chain [bay scallo 6e-27
[SCOP]         dlclm_ 1.34.1.5.11 Calmodulin [Paramecium tetraurelia 1e-15
[SCOP]         d4cln_ 1.34.1.5.10 Calmodulin [Drosophila melanogaster 2e-16
[SCOP]         dlcf_ 1.34.1.5.9 Calmodulin [African frog (Xenopus laevis) 2e-16
[SCOP]         dlahr_ 1.34.1.5.8 Calmodulin [chicken gallus gallus 4e-16
[SCOP]         d3cln_ 1.34.1.5.7 Calmodulin [rat (Rattus rattus) 2e-16
[SCOP]         dltrcb_ 1.34.1.5.6 Calmodulin [bovine (Bos taurus) 8e-08
[SCOP]         dlcll_ 1.34.1.5.5 Calmodulin [human (Homo sapiens) 2e-16
[SCOP]         dlrtpl_ 1.34.1.4.5 Parvalbumin [rat (Rattus rattus) 8e-06
[SCOP]         d5tnc_ 1.34.1.5.2 Troponin C [turkey (Meleagris gallopavo) 3e-13
[SCOP]         dlpvaa_ 1.34.1.4.3 Parvalbumin [pike (Esox lucius) 6e-06
[SCOP]         dltnp_ 1.34.1.5.1 Troponin C [chicken (Gallus gallus) 9e-11
[EC]           2.7.1.107 Diacylglycerol kinase 2e-08
[PIRKW]        blocked amino end 1e-100
[PIRKW]        phosphotransferase 2e-08
[PIRKW]        duplication 4e-17
[PIRKW]        tandem repeat 7e-06
[PIRKW]        heterodimer 4e-17
[PIRKW]        heart 6e-09
[PIRKW]        zinc 2e-08
[PIRKW]        serine/threonine-specific protein kinase 1e-06
[PIRKW]        muscle contraction 1e-08
[PIRKW]        acetylated amino end 4e-09
[PIRKW]        ATP 2e-08
[PIRKW]        skeletal muscle 6e-09
```



```

[PIRKW]      signal transduction 1e-91
[PIRKW]      protein kinase 2e-08
[PIRKW]      calcium binding 1e-100
[PIRKW]      alternative splicing 2e-13
[PIRKW]      methylated amino acid 1e-09
[PIRKW]      thin filaments 1e-08
[PIRKW]      lipoprotein 1e-101
[PIRKW]      cardiac muscle 6e-09
[PIRKW]      muscle 6e-09
[PIRKW]      myristylation 1e-100
[PIRKW]      EF hand 1e-101
[PIRKW]      retina 2e-51
[SUPFAM]     calcium-dependent protein kinase 2e-08
[SUPFAM]     unassigned calmodulin-related proteins 8e-41
[SUPFAM]     spec-related protein LpS1 7e-06
[SUPFAM]     calmodulin repeat homology 1e-101
[SUPFAM]     human diacylglycerol kinase 2e-08
[SUPFAM]     protein kinase C zinc-binding repeat homology 2e-08
[SUPFAM]     protein kinase homology 2e-08
[SUPFAM]     calmodulin 1e-101
[PROSITE]    EF_HAND 3
[PROSITE]    CK2_PHOSPHO_SITE      7
[PROSITE]    PKC_PHOSPHO_SITE      3
[PFAM]       EF_hand
[KW]         All_Alpha
[KW]         3D

```

```

SEQ      MGKQNSKL RPEVMQDLLLESTDFTEHEIQEWYKGLRDCPSGHL SMEEFKKIYGNFFPYGD
1rec-    .....HHHHHHHHHTTTTCCCHHHHHHHHHHHHHHTTTTEEEHHHHHHHHHHHTTTTC

SEQ      ASKFAEHV FRTFDANGDGTIDFREFI IALSVTSRGKLEQKLK WAFS MYDL DNGYISKAE
1rec-    HHHHHHHHHHHH-----CEEHHHHHHHHHHHHHCCCGGGHHHHHHHHHTTTTCCCEEHHH

SEQ      MLVIVQAI YKMVSSVMKMPEDESTPEKRTEKIFRQMDTNRD GKL SLEEFIRGAKSDPSIV
1rec-    HHHHHHHHHHCHCTTGGGCTTTTCHHHHHHHHHHHHCCCTTTTECHHHHHHHHHHCHHHH

SEQ      RLLQCDPSSAGQF
1rec-    HHHCCCH.....

```

Prosite for DKFzphfbr2_23b21.1

PS00005	92->95	PKC_PHOSPHO_SITE	PDOC00005
PS00005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS00006	44->48	CK2_PHOSPHO_SITE	PDOC00006
PS00006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS00006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	158->162	CK2_PHOSPHO_SITE	PDOC00006
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00018	73->86	EF_HAND	PDOC00018
PS00018	109->122	EF_HAND	PDOC00018
PS00018	157->170	EF_HAND	PDOC00018

Pfam for DKFzphfbr2_23b21.1

```

HMM_NAME      EF hand
HMM            *MFrMMDkDGDGyIDFEEFmeMMkem*
               +FR +D +GDG+IDF EF+ +++
Query          68 VFRTFDANGDGTIDFREFI IALSVT      92

30.75   100   128   1   29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
Alignment to HMM consensus:
Query          *EIqEMFrMMDkDGDGyIDFEEFmeMMkem*
               +++++F+M+D DG+GYI++ E+++++++
dkfzphfbr2     100 KLKWAFSMYDL DNGYISKAEMLVIVQAI      128

Query          176   1   29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
Alignment to HMM consensus:
HMM            *EIqEMFrMMDkDGDGyIDFEEFmeMMkem*
               +FR MD+++DG+++ EEF++ K+
Query          148 RTEKIFRQMDTNRD GKL SLEEFIRGAKSD      176

```

DKFZphfbr2_23f2

group: brain derived

DKFZphfbr2_23f2 encodes a novel 182 amino acid protein with weak similarity to S. pombe Vps29p.

No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Vps29p

complete cDNA, complete cds, EST hits
S.cerevisiae and S.pombe Vps29p are involved in vacuolar protein sorting
part of the cDNA is encoded by HSAC2350, splice pattern 4 exons

Sequenced by AGOWA

Locus: /map="12q24"

Insert length: 1016 bp

Poly A stretch at pos. 996, polyadenylation signal at pos. 974

```

1 GAATGGGGAG GAGCCAGAGG AAGAGGGCGG CGACGGTGGT GGTGACTGAG
51 CGGAGCCCGG TGACAGGATG TTGGTGTGG TATTAGGAGA TCTGCACATC
101 CCACACCCGT GCAACAGTTT GCCAGCTAAA TTCAAAAAC TCCTGGTGCC
151 AGGAAAAATT CAGCACATTC TCTGCACAGG AAACCTTTGC ACCAAAGAGA
201 GTTATGACTA CCTCAAGACT CTGGCTGGTG ATGTTTCATAT TGTGAGAGGA
251 GACTTCGATG AGAATCTGAA TTATCCAGAA CAGAAAGTTG TGAAGTTGG
301 ACAGTTCAAA ATTGGTCTGA TCCATGGACA TCAAGTTATT CCATGGGGAG
351 ATATGGCCAG CTTAGCCCTG TTGCAGAGGC AATTGATGT GGACATTCTT
401 ATCTCGGGAC ACACACACAA ATCTGAAGCA TTTGAGCATG AAAATAAATT
451 CTACATTAAT CCAGGTTCTG CCACTGGGGC ATATAATGCC TTGGAAACAA
501 ACATTATTCG ATCATTGTG TTGATGGATA TCCAGGCTTC TACAGTGCTC
551 ACCATATGTG ATCAGCTAAT TGGAGATGAT GTGAAAGTAG AACGAATCGA
601 ATACAAAAAA CCTTAAAGCC AGGCCTGTCT TGATGATTTT TGGTTTTTTT
651 TCATTGTCCT GTTGAAATCA AGTAATTAAA CATTTAAGAG CCACAAAATT
701 GTATCACTTT TATAATATTT TGCAGTAAAA TATAATACCA TCTTCTCTGT
751 TAATACATAA TTGCTCCAAG CTTCCTGTAA ACTATAAGAA TATATTAGT
801 TTACAGTATA TGGATTCTAT GAAAAAATGT CCACAACACA GTAATTGGTC
851 ACTTGTTAAG AAAAATTTAT CCTTGTAAGT ATCTTCAAAG TTGATATTTG
901 GAACTTTATT CCAAAAGTAG TGCATGTGGA GAAAGAATCT AGACTTTCTT
951 GTATACATTT TTCTCTTCTC CAGTAATAAA CAATTACCTT TCATTGAAAA
1001 AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HSAC2350 from database EMBLNEW:

Homo sapiens 12q24 PAC P424M6 Length = 167,217

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 68 bp to 613 bp; peptide length: 182

Category: similarity to known protein

Prosite motifs: RGD (60-63)

```

1 MLVLVLGLDL IPHRCNSLPA KFKLLVPGK IQHILCTGNL CTKESYDYLK
51 TLAGDVHIVR GDFDENLNP EQKVVTVGQF KIGLIHGQV IPWGDMSLA
101 LLQRQFDVDI LISGHTHKSE AFEHENKFYI NPGSATGAYN ALETNIIPSF

```

151 VLMDIOASTV VTYVYOLIGD DVKVERIEYK KP

BLASTP hits

Entry CEZK1128_6 from database TREMBL:
 "ZK1128.1"; *Caenorhabditis elegans* cosmid ZK1128
 Length = 523
 Score = 400 (140.8 bits), Expect = 2.3e-37, P = 2.3e-37
 Identities = 81/150 (54%). Positives = 106/150 (70%)

Entry S46793 from database PIR:
hypothetical protein YHR012c - yeast (*Saccharomyces cerevisiae*)
Length = 282
Score = 180 (63.4 bits), Expect = 3.7e-37, Sum P(3) = 3.7e-37
Identities = 35/71 (49%), Positives = 44/71 (61%)

Entry AB011824_1 from database TREMBL:
 "Vps29"; Schizosaccharomyces pombe mRNA for Vps29,
 partial cds. Schizosaccharomyces pombe (fission yeast)
 Length = 176
 Score = 189 (66.5 bits), Expect = 2.7e-27, Sum P(2) = 2.7e-27
 Identities = 33/72 (45%), Positives = 50/72 (69%)

Alert BLASTP hits for DKFZphfbr2 23f2, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 23f2, frame 2

Report for DKFZphfbr2 23f2.2

```

[LENGTH]      182
[MW]           20445.84
[pI]           6.29
[HOMOL]        TREMBL:CEZK1128_6 gene: "ZK1128.8"; Caenorhabditis elegans cosmid ZK1128 2e-51

[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YHR012w]
1e-27
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YHR012w] 1e-27
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YHR012w]
1e-27
[FUNCAT]       30.08 organization of golgi [S. cerevisiae, YHR012w] 1e-27
[FUNCAT]       09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YHR012w] 1e-27
[FUNCAT]       r general function prediction [M. jannaschii, MJ0623] 1e-16
[BLOCKS]       BL01269D
[BLOCKS]       BL01269A
[PROSITE]      RGD 1
[PROSITE]      MYRISTYL 4
[PROSITE]      PKC_PHOSPHO_SITE 1
[KW]           Alpha Beta

```

SEQ MLVLVLGDLHI PHRCNSLPKFKKLLVPGKIQHILCTGNLCTKESYDYLKTLAGDVHIVR
PRD cccceeeccccccccccccchhhhhhhhhhhccceeeeeeccccchhhhhhhhhhhhhccceeee

SEQ GDFDENLNYPEQKVVTVGQFKIGLIHGHQVIPWGDMA SLALLQRQFDVDILISGHTHKSE
PRD CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHCCCCCCCCCCCCC

```
SEQ      AFEHENKFYINPGSATGAYNALETNII PSFVLMDIQASTVVTVYYQLIGDDVKVERIEYK  
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCceeeeeccccceeEEEEEEcccccEEEEEEEc
```

SEQ	KP
PRD	CC

Prosite for DKF2phfbr2 23f2.2

PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	83->89	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	137->143	MYRISTYL	PDOC00008
PS00016	60->63	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2 23f2.2)

DKFZphfbr2_23124

group: intracellular transport and trafficking

DKFZphfbr2_23124.2 encodes a novel 348 amino acid protein with similarity to human glycoprotein gp36b and canine VIP36 glycoprotein.

The vesicular protein VIP36 (36 kDa vesicular integral membrane protein) shows homology to leguminous plant lectins. The protein is localized to the Golgi apparatus, endosomal and vesicular structures and the plasma membrane. VIP36 binds to sugar residues of glycosphingolipids and/or glycosylphosphatidyl-inositol anchors and might provide a link between the extracellular/luminal face of glycolipid rafts and the cytoplasmic protein segregation machinery. Gp36 is located within the endoplasmatic reticulum. For the novel protein, a lectin character is predicted. Due to the intracellular localisation of the homolog proteins, it should be involved in intracellular transport and trafficking.

The new protein can find application in modulating/blocking intracellular transport and trafficking.

strong similarity to human GP36b glycoprotein

complete cDNA, complete cds, EST hits
potential start at Bp 29 matches kozak consensua ANNatgG
similarity to lectins,

Sequenced by AGOWA

Locus: /map="2"

Insert length: 2416 bp
Poly A stretch at pos. 2394, no polyadenylation signal found

```
1  GGGGGATGAA  GGGTCGTTGG  TGGGAAAGAT  GGCGCCGACT  CTGGGACCCC
51  TTGGGTCGTG  GCAGCAGTGG  CGGCGATGTT  TGTCCGGCTCG  GGATGGGTCC
101 AGGATGTTAC  TCCTTCTTCT  TTTGTTGGGG  TCTGGGCAGG  GGCCACAGCA
151 AGTCGGGGCG  GGTCAAACGT  TCGAGTACTT  GAAACGGGAG  CACTCGCTGT
201 CGAAGCCCTA  CCAGGGTGTG  GGCACAGGCA  GTTCTCACT  GTGGAATCTG
251 ATGGGCAATG  CCATGGTGAT  GACCCAGTAT  ATCCGCCCTTA  CCCCAGATAT
301 GCAAAGTAAA  CAGGGTGCCT  TGTGGAACCG  GGTGCCATGT  TTCCTGAGAG
351 ACTGGGAGTT  GCAGGTGCAC  TTCAAAATCC  ATGGACAAGG  AAAGAAGAAT
401 CTGCATGGGG  ATGGCTTGGC  AATCTGGTAC  ACAAAGGATC  GGATGCAGCC
451 AGGGCCTGTG  TTTGGAACA  TGGACAAATT  TGTGGGGCTG  GGAGTATTTG
501 TAGACACCTA  CCCCAATGAG  GAGAAGCAGC  AAGAGCGGGT  ATTCCCTTAC
551 ATCTCAGCCA  TGGTGAACAA  CGGCTCCCTC  AGCTATGATC  ATGAGCGGGA
601 TGGCGCGCCT  ACAGAGCTGG  GAGGCTGCAC  AGCCATTGTC  CGCAATCTTC
651 ATTACGACAC  CTTCTGGTGG  ATTCTGCTACG  TCAAGAGGCA  TTTGACGATA
701 ATGATGGATA  TTGATGGCAA  GCATGAGTGG  AGGGACTGCA  TTGAAGTGCC
751 CGGAGTCCGC  CTGCCCCGCG  GCTACTACTT  CGGCACCTCC  TCCATCACTG
801 GGGATCTCTC  AGATAATCAT  GATGTCATTT  CTTGAAAGTT  GTTTGAAGTG
851 ACAGTGGAGA  GAACCCCAAG  AGAGGAAAG  CTCCATCGAG  ATGTGTTCTT
901 GCCCTCAGTG  GACAATATGA  AGCTGCCCTG  GATGACAGCT  CCACTGCCGC
951 CCTGAGTGG  CCTGGCCCTC  TTCCTCATCG  TCTTTTCTC  CCTGGTGTTC
1001 TCTGTATTGG  CCATAGTCAT  TGGTATCATA  CTCTACAACA  AATGGCAGGA
1051 ACAGAGCCGA  AAGCGCTTCT  ACTGAGCCCT  CCTGCTGCCA  CCACTTTTGT
1101 GACTGTCAAC  CATGAGGTAT  GGAAGGAGCG  GGCCTGGCC  TGAGCATGCA
1151 GCCTGGAGAG  TGTCTTGTG  TCTAGCAGCT  GGTGGGGGAC  TATATTCTGT
1201 CACTGGAGTT  TTGAATGCAG  GGACCCCGCA  TTCCCATGGT  TGTGCATGGG
1251 GACATCTAAC  TCTGGTCTGG  GAAGCCACCC  ACCCCAGGGC  AATGCTGCTG
1301 TGATGTGCCT  TTCCTGCAG  TCCTTCCATG  TGGGAGCAGA  GGTGTGAAGA
1351 GAATTTACGT  GGTGTGATG  CCAAAATCAC  GGAACAGAAT  TTCATAGCCC
1401 AGGCTGCCGT  GTTGTGTTG  TCAGAAGGCC  CTTCTACTTC  AGTTTGAAT
1451 CCACAAAGAA  TAAAAACTG  GTAACACCAC  AGGCTTTCTG  ACCATCCATT
1501 CGTTGGGTTT  TGCAATTTG  CCAACCCCTC  GCCTACCTGA  GGAGCTTTCT
1551 TTGAAACCA  GGATGGAAC  TTCTTCCCTG  CCTTACCTTC  CTTTCACTCC
1601 ATTCATTGTC  CTCTCTGTG  GCAACCTGAG  CTGGGAAAGG  CATTTGGATG
1651 CCTCTCTGTT  GGGGCTGGG  GCTGCAGAAC  ACACCTGCGT  TTCGCTGGCC
1701 TTCAATAGGT  GGGCCTAGG  AGATGGCTTT  CTGCTTTGGA  TCACTGTTCC
1751 CTAGCATGGG  TCTTGGGTCT  ATTGGCATGT  CCATGGCCTT  CCAATCAAG
1801 TCTCTTCAGG  CCCTCAGTGA  AGTTTGGCTA  AAGGTTGGTG  TAAAAATCAA
1851 GAGAAGCCTG  GAAGACACCA  TGGATGCCAT  GGATTAGCTG  TGCAACTGAC
1901 CAGCTCCAGG  TTTGATCAAA  CCAAAAGCAA  CATTGTTCAT  GTGGTCTGAC
1951 CATGTGGAGA  TGTTTCTGGA  CTTGCTAGAG  CCTGCTTAGC  TGCATGTTTT
2001 GTAGTTACGA  TTTTGGAAAT  CCCTCTTTGA  GTGCTGAAAG  TGTAAGGAAG
2051 CTTTCTTCTT  ACACCTTGGG  CTTGGATATT  GCCCAGAGAA  GAAATTTGGC
2101 TTTTTTTTCT  TAATGGACAA  GGGACAGTTG  CTGTTCTCAT  GTTCCAAGTC
2151 TGAGAGCAAC  AGACCCCTCAT  CATCTGTGCC  TGGAAGAGTT  CACTGTCAAT
2201 GAGCAGCACA  GCCTGAGTGC  TGGCCTCTGT  CAACCCCTAT  TCCACTGCCT
```

2251 TATTTGACAA GGGGTTACAT GCTGCTCACC TTACTGCCCT GGGATTAAAT
 2301 CAGTTACAGG CCAGAGTCTC CTTGGAGGGC CTGGAACCTCT GAGTCCTCCT
 2351 ATGAACCTCT GTAGCCTAAA TGAAATTCTT AAAATCACCG ATGGAACCAA
 2401 AAAAAAAAAA AAAAAA

BLAST Results

Entry HS622145 from database EMBL:
 human STS WI-6746.
 Score = 1079, P = 5.1e-43, identities = 219/223

Entry G42541 from database EMBLNEW:
 SHGC-58649 Human Homo sapiens STS genomic, sequence tagged site.
 Score = 1091, P = 1.7e-43, identities = 219/220

Medline entries

94265253:
 A putative novel class of animal lectins in the secretory pathway
 homologous to leguminous
 lectins.

94208543:
 VIP36, a novel component of glycolipid rafts and exocytic carrier
 vesicles in epithelial cells.

Peptide information for frame 2

ORF from 29 bp to 1072 bp; peptide length: 348
 Category: strong similarity to known protein

1 MAATLGPLGS WQQWRRCLSA RDGSRMLLLL LLLGSGQGPGQ QVGAGQTFEY
 51 LKREHSLSKP YQVGTGSSS LWNLMGNAMV MTQYIRLTPD MQSKQALWN
 101 RVPCFLRDWE LQVHFKEHQ GKKNLHGDGL AIWYTKDRMQ PGPVFGNMDK
 151 FVGLGVFVDT YPNEEKQQR VFPYISAMVN NGSLSYDHER DGRPTLGGC
 201 TAIVRNLHYD TFLVIRYVKR HLTIMMDIDG KHEWRDCIEV PGRVLRPGYY
 251 FGTSSITGDL SDNHDVISLK LFELTVERTP EEEKLHRDVF LPSVDNMKLP
 301 EMTAPLPPLS GLALFLIVFF SLVFSVFAIV IGIIYLNKQW EQSRKRFY

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23124, frame 2

PIR:G01447 GP36b glycoprotein - human, N = 1, Score = 1001, P =
 5.9e-101

SWISSPROT:VP36 CANFA VESICULAR INTEGRAL-MEMBRANE PROTEIN VIP36
 PRECURSOR (VIP36)., N = 1, Score = 990, P = 8.6e-100

TREMBL:CET04G9_2 gene: "T04G9.3"; Caenorhabditis elegans cosmid
 T04G9., N = 1, Score = 614, P = 6e-60

PIR:S42626 ER-golgi intermediate compartment protein - human, N = 2,
 Score = 397, P = 1e-42

>PIR:G01447 GP36b glycoprotein - human
 Length = 356

HSPs:

Score = 1001 (150.2 bits), Expect = 5.9e-101, P = 5.9e-101
 Identities = 197/356 (55%), Positives = 256/356 (71%)

Query: 1 MAATLGPLGSWQQWRRCLSA RDG-----SRMLLLLLLLGSGQGPGQVVGAGQTFEYLK 52
 MAA G + W RRCL R G + L LLLLLGS + G + E+LK
 Sbjct: 1 MAAE-GWIWRWGWGRCLG-RPGLLGPGPGPTPLFLLLLLGSVTA--DITDGNS-EHLK 55

Query: 53 REHSLSKPYQGVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQGALWNRVPCFLRDWELQ 112
 REHSL KPYQGVG+ S LW+ G+ M+ +QY+RLTPD +SK+G++WN PCFL+DWE+
 Sbjct: 56 REHSLIKPYQGVGSSSMPLWDFQGSTMLTSQYVRLTPDERSKEGSIWNHQPCLKDWEMH 115

Query: 113 VHFKIHGQGGKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFGVGLGVFVDTPNEEKQQERV 172
 VHFK+HG GKKNLHGDG+A+WYT+DR+ PGPVFG+ D F GL +F+DTYPN+E ERVF
 Sbjct: 116 VHFKVHGTGKKNLHGDGIALWYTRDLVPGPVFGSKDNEHGLAIFLDTPNDETT-ERVF 174

Query: 173 PYISAMVNNGSLSYDHERDGRPTLGGCTAIVRNLYDTFLVIRYVVKRHLTIMMDIDGKH 232
 PYIS MVNNGSLSYDH +DGR TEL GCTA RN +DTFL +RY + LT+M D++ K+
 Sbjct: 175 PYISVMVNNGSLSYDHSKDGRTLGGCTADFRNRDHTFLAVRYSRGLTVMTDLEDKN 234

Query: 233 EWRDCIEVPGVRLPRGYFGTSSITGDLSDNHDVISLKLFEVTVERTPEEEKLHRDVFLP 292
 EW++CI++ GVRLP GYFYG S+ TGDLSNHD+IS+KLF+L VE TP+EE + P
 Sbjct: 235 EWKNCIDITGVRLPTGYFYGASAGTGDLSNHDIIISMKLFQLMVEHTPDEESIDWTKIEP 294

Query: 293 SVDNMKLPMTAPLP-----PLSGLALFLIVFFSLVFSVFAIVIGIILYNKWQEQSRK 345
 SV+ +K P+ P PL+G +FL++ +L+ V V+G +++ K QE++ K
 Sbjct: 295 SVNFLKSPKDNVDDPTGNFRSGPLTGWRVFLLLCALLGIVVCAVVGAVVFQKRQERN-K 353

Query: 346 RFY 348
 RFY
 Sbjct: 354 RFY 356

Pedant information for DKFZphfbr2_23124, frame 2

Report for DKFZphfbr2_23124.2

[LENGTH] 348
 [MW] 39711.10
 [PI] 8.55
 [HOMOL] PIR:G01447 GP36b glycoprotein - human le-101
 [PIRKW] lectin 2e-37
 [PIRKW] transmembrane protein 2e-37
 [PIRKW] endoplasmic reticulum 2e-37
 [PIRKW] Golgi apparatus 2e-37
 [PROSITE] AMIDATION 1
 [PROSITE] MYRISTYL 5
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 3
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] Alpha_Beta
 [KW] SIGNAL PEPTIDE 39
 [KW] LOW_COMPLEXITY 7.76 %

SEQ MAATLGPLGSWQWRRCLSDRGSRMLLLLLLLGSGQGPQVAGQTFEYLKREHSLSKP
 SEGxxxxxxx.....
 PRD cccccccccccccccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhccc

SEQ YQGVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQGALWNRVPCFLRDWELQVHFKIHGQ
 SEG
 PRD cccccccccceccccccccceeeccchhhhhccccccccccchhhhhhhheeeec

SEQ GKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFGVGLGVFVDTPNEEKQQERVFPYISAMVN
 SEG
 PRD cccccccccceeeccccccccccccccccccccceeeccccccccccccccccceec

SEQ NGSLSYDHERDGRPTLGGCTAIVRNLYDTFLVIRYVVKRHLTIMMDIDGKHEWRDCIEV
 SEG
 PRD cccccccccccccccccccccccccccccceeehhhhhhheeecccccccccccccc

SEQ PGVRLPRGYFGTSSITGDLSDNHDVISLKLFEVTVERTPEEEKLHRDVFLPSVDNMKLP
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccc

SEQ EMTAPLPPLSGLALFLIVFFSLVFSVFAIVIGIILYNKWQEQSRKRFY
 SEGxxxxxxx.....
 PRD cccccccccchhhccc

Prosite for DKFZphfbr2_23124.2

PS00001	181->185	ASN_GLYCOSYLATION	PDOC00001
PS00002	35->39	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	19->22	PKC_PHOSPHO_SITE	PDOC00005

PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDOC00006
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	96->102	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00009	120->124	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_23124.2)

DKF2phfbr2_23n16

group: signal transduction

DKF2phfbr2_23n16.1 encodes a novel 292 amino acid protein with weak similarity to putative phosphatidylinositol-4-phosphate 5-kinase of *Arabidopsis thaliana*.

The novel proteins contains a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore the new protein should be involved in intracellular signal transduction.

The new protein can find application in modulating/blocking intracellular signal transduction pathways.

similarity to putative phosphatidylinositol-4-phosphate 5-kinase

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2936 bp

Poly A stretch at pos. 2916, polyadenylation signal at pos. 2873

```
1  GGGGGCGCTC  CCGAGAAAGA  GTGAGGGCGC  GACGCGCACC  AACGGTGGAG
51  GGATGTTTCA  GCAGCCCTTG  AGAAGGAAGA  GGAGGAAGCT  GAGGGCCCCG
101 TGAGGGCGCA  GGACCTGAGG  GAGTCCTACA  TCCAGCTCGT  CCAGGGTGTG
151 CAGGAGTGGC  AGGATGGTTG  CATGTACCAG  GGGGAGTTTG  GGTGAACAT
201 GAAGCTTGGG  TATGGCAAA  TCTCTTGGCC  CACAGGCGAG  TCATACCATG
251 GGCAGTTTFA  CCGGGACCAC  TGCCATGGCC  TGGGTACCTA  CATGTGGCCA
301 GATGGCTCCA  GTTTCACGGG  CACATTTTAC  CTCAGCCACC  GAGAAGGCTA
351 CGGCACCATG  TACATGAAGA  CACGGCTTTT  CCAGACTCAC  TGCCACAACG
401 ACATTGTCAA  CCTTCTCCTG  GACTGTGGGG  CCGACGTGAA  CAAGTGCTCA
451 GATGAGGGTC  TCACGGCACT  CAGCATGTGT  TTCCTCCTCC  ACTACCCCGC
501 CCAGTCTCTC  AAGCCCAATG  TTGCTGAACG  GACCATACCT  GAGCCCCAGG
551 AACCTCCAAA  ATTCCAGTT  GTTCCAATCC  TTTTCATCAT  ATTTATGGAC
601 ACAAACTGGG  AGTCTCTGTA  CTATGAGGTG  AACGTGCCTT  CCCAGGGTAG
651 CTATGAGCTG  AGGCCACCGC  CAGCACCCT  GCTCCTGCCA  CGCGTCTCAG
701 GCAGCCACGA  GGGCGGCCAC  TTCCAGGACA  CCGGGCAGTG  TGGGGGGTCC
751 ATAGACCACA  GGAGCAGCTC  TCTGAAGGGG  GACTCCCCCT  TGGTGAAGGG
801 CAGCCTTGGC  CATGTGGAAA  GCGGGCTTGA  GGACGTGTTG  GGAGACACAG
851 ACCGGGGCAG  TCTGTGCAGT  GCTGAGACGA  AATTTGAGTC  CAACTTGTGT
901 GTGTGCGACT  TCTCCATCGA  GCTCTCGCAG  GCCATGCTGG  AGAGAAGCGC
951 CCAGTCCAC  AGCTTGCTGA  AGATGGCCTC  GCCCTCACCG  TGCACCAGCA
1001 GCTTCGACAA  AGGGACCATG  CGGAGGATGG  CGCTGTCCAT  GATCGAGTAG
1051 GTCCTGGCAC  CAGCTGGTGG  GGGTGGAGGG  CCACCATCAG  GGCTGAATCC
1101 TATGCTCAGC  AGACCCACGT  CTCTTCCCTG  TGCCAGTGGG  AGGCGTTGTG
1151 TCTGGAGATG  TGTGTCTGAA  TGTGTGAGCA  TCCCTGTGTC  GGTGGCTCCA
1201 TGCCATGGCC  AGCCCTGTGG  GGGTGGCCAG  GTGACGGGCT  GTTTTCAGTG
1251 CCACCCACGC  CCTGTGGGGG  TGCCACGGTG  ACGGGCTGTT  TTCAGTACCA
1301 CGCCAGCCCT  GCTTTGGCCT  TTGGCACTGG  CCTGAAGTGT  CTCTGTGGGA
1351 GCCTCAGCAG  GGGCCACTGT  CAGGGGTCTT  ATCCTAGCCA  TAGTGCACGT
1401 GAGTGACACC  TGCCTGGGCA  GCTCTCACAC  CCCTGCTGTC  CACCTGTGCT
1451 ATACCACTGT  GTCTCAAAAT  GTGGTCTATG  CACCCCGGGT  GGTCCAAGAC
1501 CCTTTTCAGG  AGTCTGTGGG  GTCAAAATGA  TTCTCTTGAT  AACCTTGAGA
1551 CTCTGTAGC  CTTCTCCTTG  TGTGATGTT  GGTGGATGGT  ATGAAGACAG
1601 GGCCGTGACG  ACCACAGCC  CCCAGCGTGC  AGGGCAGCAG  TGCCCGGCTT
1651 GCTTGGGGGC  ATGGTATTCC  TTCACCACGG  TGTGCACTTG  CGGGGATGCC
1701 TGTCTCACTG  AAGAAATGCT  TTGACTAAGC  AGAAAAGCAA  TGACAAATTG
1751 CATTTAAATC  TGCTCCTTGC  GTACACACCC  CTCGAATATT  CTGGGTCGGA
1801 AAACATGGGA  AGGACACTGA  TGTGTGTCTG  CCACAGACCA  AGGCACACCG
1851 CTTCCCCGCA  AGAAGCGCTT  CCCCCAGGGC  CAGAGTAGCA  ACAGAATGCG
1901 GCATCTTCCC  AACCTCCTGC  CCCATTTTTC  ATTGGAAGAA  TGACCACTGG
1951 TATGTGGCTG  TTCATTCTCC  TGAACACAGC  CTGCCACTTT  AAGGAAAACA
2001 TATGACACTA  TTTGTTGCTG  GCGAAATTTA  CATTTTCAAG  TGAATAGCAG
2051 AATTCTGGAC  ACTTGCCACC  ACCACCAAAA  CCTTCATAGC  TTCCCTTAAC
2101 TTTGAGACAT  GGGTGTTCAG  AGGTTTTTCA  CGTGAGATGG  CGTTAGCAGC
2151 GCAGTTTTGT  GATACTGCCT  GAAGACATGC  CGACAGTGCC  CAGATCTCTT
```



```

2201 CTATTGGTGA GCCAGCTTTT CCCACACGGC CAAGTCTCTGA TGTGGAACCA
2251 TTGCCAGGTG GGTGAAGATC CATTGACAGT GAGAGGTGGG CCCGTGGGCT
2301 TCAGTGCAGC CAGGCGCAGA AGGCTGGTTC ATGAGTGTCC AGCTCCGCCA
2351 GGTAGCTAGC TCACCACCCC CAGCCTGGGT TCATGTAGTT CAAATAGGAA
2401 GACCACGATG ATCAGAAAAGG CTGCTCAAAT ACTCCTTCGT CCAGCCGCGT
2451 ACCTGGGGGA GGCTGAATCT CCACTCACTT CCACCAAGGC TGTGCAGAGC
2501 AGATAGGGGA ATCCAGCAAA GGTGGAAAAC AGTGCCATCC TTCTCCCCAA
2551 CTGGTTTTGT TTTGTAAAAT AACTTTTTGT GACAGTGTTA CTTATTAGTA
2601 ACATGCAGTG GGTTTGTTAT GGTTAACAAG TTGGTGAGCA TTATTGAGAG
2651 GTGAAGCCAG CTGAGCTTCT GGGTTGGGTG GGGACTTGA GAACTTTTGT
2701 GTCTAGCTAA AGGATTGTAA ATGCACCAAT CAATGCTCAG TGTCTAGCTA
2751 AAGGATTGTA AATGCACCAA TCAGCACTCT GTAAAATTGA CCAATCAGCG
2801 TTCTGTAAAA TGGACCAATC AGTGGTCTGT AAAATGGACC AGTCAGCAGG
2851 ATGTGGGCGG GGCCAAAAAA GGAATAAAAA GCTGGCCACC GCCAGGCTCC
2901 CCACCAGCCT GCAGCGAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 172 bp to 1047 bp; peptide length: 292
 Category: similarity to unknown protein
 Prosite motifs: WW_DOMAIN_1 (19-24)

```

1 MYQGEFGLNM KLGYGKFSWP TGESYHGQFY RDHCHGLGTY MWPDGSSFTG
51 TFYLSHREGY GTMYMKTRLF QTHCHNDIVN LLLDCGADVN KCSDEGLTAL
101 SMCFLHLYPA QSFKNVAER TIPEPQEPK FPVVVILSSS FMDTNLESly
151 YEVNVPSQGS YELRPPAPL LLPRVSGSHE GGHFQDTGQC GGSIDHRSSS
201 LKGDSPLVKG SLGHVESGLE DVLGDTDRGS LCSAETKFES NLCVCDFSIE
251 LSQAMLERSA QSHSLKMAS PSPCTSSFDK GTMRRMALSM IE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23n16, frame 1

TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for
 AtPIP5K1, complete cds., N = 2, Score = 138, P = 1.1e-06

TREMBL:AF019380_1 product: "putative phosphatidylinositol-4-phosphate
 5-kinase"; Arabidopsis thaliana putative
 phosphatidylinositol-4-phosphate 5-kinase mRNA, complete cds., N = 2,
 Score = 138, P = 1.4e-06

PIR:T02098 probable phosphatidylinositol-4-phosphate 5-kinase -
 Arabidopsis thaliana, N = 2, Score = 135, P = 6.7e-06

>TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for
 AtPIP5K1, complete cds.
 Length = 683

HSPs:

Score = 138 (20.7 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
 Identities = 23/61 (37%), Positives = 35/61 (57%)

```

Query:      1 MYQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTTFYLSHREGY 60
              MY+G++      G GKFSWP+G +Y G+F      G GT+      DG ++ GT+      + G+
Sbjct:     34 MYEGDWKRGKASGKGKFSWPSGATYEGEFKSGRMEGFGFTTGADGDTYRGTWVADRKHGH 93

Query:      61 G 61
              G
Sbjct:     94 G 94

```

Score = 112 (16.8 bits), Expect = 9.7e-04, Sum P(2) = 9.7e-04
Identities = 19/51 (37%), Positives = 27/51 (52%)

Query: 12 LGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGYGT 62
+G GK+ W G Y G + R G G + WP G+++ G F EG+GT
Sbjct: 22 IGSGKYLWKDGCMEYEGDWKRGKASGKGKFSWPSGATYEGEFKSGRMEGFGT 72

Score = 97 (14.6 bits), Expect = 4.4e-02, Sum P(2) = 4.3e-02
Identities = 19/60 (31%), Positives = 32/60 (53%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGYGT 61
Y+GEF G+G F+ G++Y G+ D HG G + +G + GT+ + ++G G
Sbjct: 58 YEGEFKSGRMEGFGTFTGADGDTYRGTVVADRKHGHGQKRYANGDFYEGTWRRNLQDGRG 117

Score = 93 (14.0 bits), Expect = 1.2e-01, Sum P(2) = 1.1e-01
Identities = 18/62 (29%), Positives = 34/62 (54%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGYGT 61
Y+G + + K G+G+ + G+ Y G + R+ G G Y+W +G+ +TG + + G G
Sbjct: 81 YRGTVVADRKHGHGQKRYANGDFYEGTWRRNLQDGRGRYVVRNGNQYTGEWRIGVISGKG 140

Query: 62 TM 63
+
Sbjct: 141 LL 142

Score = 91 (13.7 bits), Expect = 2.0e-01, Sum P(2) = 1.8e-01
Identities = 18/51 (35%), Positives = 24/51 (47%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGYGT 61
Y GE+ + + G G WP G Y G + G G + W DGSS G +
Sbjct: 127 YTEWRIGVISGKGLLVWPNGNRYEGLWENGIPKGNVFTWSDGSSCVGAW 177

Score = 90 (13.5 bits), Expect = 2.6e-01, Sum P(2) = 2.3e-01
Identities = 17/60 (28%), Positives = 31/60 (51%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGYGT 61
Y+G + N++ G G++ W G Y G++ G G +WP+G+ + G + +G G
Sbjct: 104 YEGTWRRNLQDGRGRYVVRNGNQYTGEWRIGVISGKGLLVWPNGNRYEGLWENGIPKGN 163

Score = 45 (6.8 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
Identities = 14/62 (22%), Positives = 26/62 (41%)

Query: 215 VESGLEDLVLDTRGSLCSAETKFESNLCVDCF--SIELSQAMLSAQSHSLKMASPS 272
V+SG + G+ +C E+ E+ CD ++E S +R + + +
Sbjct: 205 VDSGAGSLGGEKVFPRIWESDGEAGDITCDIIDNVEASMIYRDRISVDRDGRFQFKKN 264

Query: 273 PC 274
PC
Sbjct: 265 PC 266

Pedant information for DKFZphfbr2_23n16, frame 1

Report for DKFZphfbr2_23n16.1

[LENGTH] 292
[MW] 32214.44
[pI] 5.51
[HOMOL] TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for AtPIP5K1,
complete cds. 7e-08
[BLOCKS] BL01137A Hypothetical YBL055c/yjjv family proteins
[PROSITE] WW DOMAIN_1 1
[PROSITE] MYRISTYL 5
[PROSITE] CK2_PHOSPHO_SITE 7
[PROSITE] PKC_PHOSPHO_SITE 5
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 4.11 %

SEQ MYQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGY
SEG
PRD ccc
SEQ GTMYMKTRLRFQTHCHNDIVNLLDCGADVKNKCSDEGLTALSMCFLHYPASFKPNVAER
SEG
PRD cccchhhhhheccccchhhhhccccccccccccccccchhhhhhhhhcccccccccccc
SEQ TIPEPQEPKFPVVPILSSSFMDTNLESLEYEVNVPSQGSYELRPPAPLLLPRVSGSHE

```

SEG .....XXXXXXXXXXXX.....
PRD eccccccccceeeeeccccccccceeeeeccccccccccccccccccccccccc

SEQ GGHFQDTGQCGGSIDHRSSSLKGDSPLVKGS LGHVESGLEVDV LGD TD RGS L C S A E T K F E S
SEG .....
PRD cccccccccccccccccccccceccccccccccccccccccccceeeeecccc

SEQ NLCVCDFSIELSQAM LERSAQSHSLKMASPSPCTSSFDKGTMR RMALSMIE
SEG .....
PRD cccccchhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhccc

```

Prosites for DKFZphfbr2_23n16.1

PS00005	55->58	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	226->229	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	121->125	CK2_PHOSPHO_SITE	PDOC00006
PS00006	140->144	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	217->221	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00008	45->51	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	177->183	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	229->235	MYRISTYL	PDOC00008
PS01159	19->44	WW_DOMAIN_1	PDOC50020

(No Pfam data available for DKFZphfbr2_23n16.1)

DKFZphfbr2_23o24

group: brain derived

DKFZphfbr2_23o24 encodes a novel 139 amino acid protein with similarity to CAAX-box proteins.

The CAAX box is a prenyl group binding site found in a number of eukaryotic proteins, such as which is found in Ras- and ras-like proteins such as Rho, Rab, Rac, Ral, and Rap, as well as in nuclear lamins A and B, some G protein alpha and gamma subunits and some dnaJ-like proteins. These proteins are posttranslationally modified at this site by the attachment of either a farnesyl or a geranyl-geranyl group to a cysteine residue.

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to lectins

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3564 bp

Poly A stretch at pos. 3541, no polyadenylation signal found

```
1 GAATGGCTCC GCAGATGGCC GGCAGTGA GAAGCAAGA AGCGGAGGAG
51 ATGGGCCCTTC AGCAGGGGGT TGGGGGGGA GCTTTAACT GAGCCCTGTA
101 AACATGGCAG AACTGCTCAG TGGGAGACTC TCAGCACAGA CGGTCATGGG
151 GAAGTGAGTG CAGTTCATTT GTAATCTTGT TGTCCAGTTC TGGGTTTTTT
201 TTGTTTGTTC CGTAACTTTA AAGGTATGCA CTTTATATAG ATTTATTTAT
251 TTGCTGGGAC CGTTACTCAG AGTTCCTAGA AATGTACACA GCTTTTTTAC
301 CAGGGTTACT CCTCAGAATC ACTTGTCAC TCTTTAAATG AATGAATGAA
351 TGTGCCAGGC CCTATGCCGT GAGGTGGGA GCTTCATCTA CATCACATTC
401 TAACAGGTGA CCACTGGGGT AAGCACTGTG TGAATGCAA GCCAGGGTGT
451 GTTTCATCA ACACCCAGAT GACCGTGCTT ATGTGCCCTT GTTGTCTTCC
501 CTCCAGGACT GCCTCCTCAC CCCACCCCTT TCTGCAGCTC CTCATCTAAA
551 CATCTCGCCT GGTGAGGTCA CGGCTTAGCC TGTGGCCAG TGGCCCCACC
601 ACCATCCTTC CCCCTGTGCA GATTGGAGGA GGCCAGGTCT CTCCCCTTAG
651 CTCTATGTGC CCCTTCACCC CCCATGGCAC AGATGAGACA TTCACAGAGT
701 TTGCAGATGA TGAAGAGAA GACTCCAGGT TGCCAGGTGT GTCCACTCTC
751 AGGAACCCCC AGCCCAAGCC TCACTGCTCG TGTCCCAGC CAACCCAGC
801 ACGGGGGATA CGCCGGTGTG GTTTCCTGTC TCAGATACAA CCAGTTACCA
851 GAAACGACCT CACCCCTCCA ACCACTTTCC AAGGTGCCAG GACAGAGAAG
901 CCCTTCACTG GCCCACCAG GGCAGTTGAC AGAGGGATGC CCTCTTGA
951 GGGGAGCCTC ACCTTACCC ACAGGGCCGC GGCCTTGTC TGGATTCTCA
1001 CCGGGGCAGT CACGTACAGG TGGAGAGGTC CCATGTCAGC CAGTCTTTG
1051 GTGGGGTCA TGTAGTCTGA AATGACCTGC CGATGGTCCA GGCTGAGCCA
1101 GGAAGCTGA GCCTGGGTGC CTTTGTGTG CTTACTCTGA CTTGAGTTGG
1151 ATTCAATGCA CAGACCCACC TTCTTGAGCA ACAACACATA TAGCCACCAA
1201 CACAAGAGCC AGGCACACAC TGAGCAGAGA AAGTCCCTGT CGCCTCACCA
1251 CCCAAAACCT CCAGCTTTGC AGAGACCAAG GTTCTTCTCT ACCTTTGCAG
1301 AAGCCTCTGT GACCAAAACC GGAGCTTGCC CTTCTGAGGC CTCTAGCATT
1351 TCTCCAGGTG TTTTTCAGAG GACTTGGTTT AAATTTGTTC ACCCCAAATG
1401 TTGTCTTTCC CGGATCATGA AAGGATCTGC CGCAAAGGTG AATCTGAGTC
1451 TCCTCAGAGT CATATGAGAC TGAAGCTGCT TATAACATTT CCGTGACCTA
1501 ATAAGTCTTC CAAAATGTA GGGTATTAAAG AGTTTAGTGA CATTAAAAAG
1551 TTTAGTCGAA AATATCGTGA TTCAGGTATA TTTAGACATT TGATTCAATG
1601 CAAATTGCCA CTGTTAAGC AAAACACACC CCAAGCATAT TAATGCCTAG
1651 ATATTTCAAA CCCTTTCTG CCCACACATT CTTAAAAATA ATATACTGAG
1701 AAATCTATAT ACAGGTTTTT TTTTAATTAG CTTGGAAAAG AGCAGTTGTA
1751 TTCTGTTTGA ACAGCTGCTA ATGTCAATTC CTGTGGGAAG AAAGACCAAA
1801 GAACATGGAG TTACACCAAG AATTTTAAAA CAAAGACGCT GTCCCTTTCC
1851 TGAGCACCGT GCAGCCAAGA CTGAGAGATC AGTCTGAGAC CTGTGATTAA
1901 GGAGTGTTTT CTACATAGCG TATAATTATG GAGCCACACA AGTGGGCCAT
1951 TACTCTGTGT AGTGCTTCAT GTTTGAGGTA TTTTCGTGTT CCAACTTACA
2001 TTAAGTGTTT TATAAACAG GAAAAATCCA CGAGCAGGTA TTGACACTAT
2051 CCATATTAGA TCATCACAAA ATTATATATA TAGCAGAGTC ATAAACAATG
2101 AGAAACGGTC TTCCACACT TGCTTTAAAT GGCCATGACC TAGTGTTTAG
2151 GGAAGCAGT AAAATCAGCG AGGAGCTCGT GGGAAAAATG AGACGGGGCC
2201 TGAGGGGGTG ACTCATGGGC CAAGCAGGGC CACACAGGTA CCAGGCCGCC
2251 ACCTCCTCTC CTGCTCTCA CTCTCTGGAG ACTGGAATTC CTTTACTGCC
2301 TCCTTTCTGA CATTTCTTAG ACATCAGACT TTGCTACTTA GTACACAAAC
2351 GGGGTTCCCT TTTAAATTG TTAATCTAG TTAGCATTG CAGAAGCTGT
2401 GAAAAATTAC AGAGAGATGA TGTGTGGGT AAGAGATGGT TTAAGTCTC
```

```

2451 AGCTTGCTGT TTTTCATTAA GTGTCTTGAA AATGAGTAAG TGGCGTTCCT
2501 GGAGGGGAAC AATCATATAA TTCCGCAGGG TGGGTCTAAA CTGTGTTTCT
2551 GATAGTGTTT AGCAGCTCAT GGCTCTGAGG GCACCTGATA ACACAGCAGC
2601 CAGGCGCTGA TGAGAAGTGT GTGCCAGACA GACCCGAGTG TGGCTTGGCT
2651 CTTGCCTTAT GTTCCTTTCT CTGTTTCAGAG AAGCGTGAGA TGAGATTTTG
2701 TGATTATATT GCACTCCTTG GGCTGACTTT CCCATGCACA GAATGTTTTA
2751 CACATCCTGA TAGCTGAGCT GAAAAATGCAA AGAGAAGGGA AAATGCCTTA
2801 AATTGTTCTG GCTAATTTAG AAGCAGCAGG CCTTGGAAGT CTTTGTCCCTG
2851 TGTCCCTGAA CAAATCTTAT GGGAGCTCTG GTACCTATGC CAGAAAATGC
2901 ACATAGGCAC AACACTTTTA CATAACGTT CACACACCCC ACCCTTATGG
2951 AGAACTTTTT TCTAAATAAG AGAAAGAAAA ATTTTAAAGC TTACAAGTTA
3001 TGTTTAGTGA TTTTACATGG TTCAGAAAAA AAGACATGAA GCGGTATAAA
3051 CTGAGAAGTC TTGTTCCAC AACCCACGT GCCAGGTACA CATAACCATT
3101 TTTATTCACC TCTAGCTTGT GCTTCCAATG TTTGTTAGGC ATATGTAAAT
3151 AAGTGAATAG ATAAGCATT CTCCCTCCTT TTGCTGACAT GAGTGGTGGC
3201 ATGTTTTGCC CCTGGCTTTT ATCCCTTGAC CCCATCCAG TACCTAGAGA
3251 CCTGCTTCAT TTTTTAGAT GTGTAATACT TCATGTGTGC GTGTGCCTTA
3301 GTGATTAACT CGTGCACGT GCAGGGACAT CGGGCTGGGA TCAGTTTGT
3351 CACTGATATA TACAGCGCTG CGGGAGATAC CCTCACATGT GTATCATTTG
3401 GTCCATGTGC AGGTGTGTCT GGAAGATAGA ATTCTAGGCG TAGAATTGAT
3451 AGGTAAATG TATTATAGG GAAAAAATCA ATATAAACT TTGCGTGTA
3501 TGATATTGCG GTGCTTTTTT TTTTAATTTT TTTACCCAAA TAGTAAAAAA
3551 AAAAAAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 656 bp to 1072 bp; peptide length: 139
 Category: similarity to known protein

```

1 MSPSPPMQAM RHSQSLQMM EKTGPGQVCP LSGTPSPSLT ARVPSQPQHG
51 GYAGAVSLLR YNQLPETTSP LQPLSKVPGQ RSPSLAHFPGQ LTEGCPPWRG
101 ASPLPTGPRP CPGFSPGQSR QDGEVPCQPV LWWGSCSLK

```

BLASTP hits

Entry CEEGAP7.1 from database TREMBL:
 gene: "EGAP7.1"; Caenorhabditis elegans cosmid EGAP7.
 Score = 123, P = 2.3e-07, identities = 35/103, positives = 44/103

Entry MMBPC35.1 from database TREMBL:
 Mouse carbohydrate binding protein 35 mRNA, 3' end.
 Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Entry A28651 from database PIR:
 galactose-specific lectin - mouse >TREMBL:MMMAC2A.1 Mouse mRNA for
 Mac-2 antigen
 Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Alert BLASTP hits for DKFZphfbr2_23o24, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23o24, frame 2

Report for DKFZphfbr2_23o24.2

```

[LENGTH]      139
[MW]           14748.91
[pI]           8.90
[PROSITE]     PRENYLATION    1

```


DKFZphfbr2_23o5

group: brain derived

DKFZphfbr2_23o5 encodes a novel 360 amino acid protein with no known similarity

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

potential start at Bp 24 matches Kozak consensus ANNatgG

Sequenced by AGOWA

Locus: /map="7q21-q22"

Insert length: 1736 bp

Poly A stretch at pos. 1714, polyadenylation signal at pos. 1680

```
1 GGGGGAGGAT CAAAGTAGGC AAGATGGCGT CGAGCGGCGG GGAGCCAGGG
51 AGTTTATTG ATCACCACGT CCAGAGGGCG GTATGCGACA CACGGGCCAA
101 ATATCGAGAG GGACGACGGC CTCGTGCTGT GAAGGTATAT ACAATCAATT
151 TGGAATCTCA GTACTTATTA ATACAAGGAG TTCCTGCTGT GGGAGTCATG
201 AAGGAATTAG TTGAGCGATT CGCTTTATAT GGTGCAATTG AACAGTACAA
251 TGCTCTAGAT GAATACCCAG CAGAAGACTT TACTGAAGTT TATCTTATTA
301 AATTATGAA CTTACAAAGT GCAAGGACAG CCAAGAGAAA AATGGATGAA
351 CAGAGTTTCT TCGGTGGATT GCTTCATGTG TGCTATGCTC CAGAAATTGA
401 AACAGTTGAA GAAACTAGAA AAAAAGTACA AATGCGGAAG GCATATGTAG
451 TAAAAACTAC TGAAAAATAA GACCATTACG TGACAAAGAA GAAATTGGTT
501 ACAGAGCATA AAGACACAGA GGATTTTAGA CAAGACTTCC ACTCAGAGAT
551 GTCTGGATTT TGTAAGCTG CTTTGAACAC TTCTGCGAGG AACTCAAATC
601 CTTATCTTCC GTATTCCTGT GAATTGCCTT TATGTTATTT CTCCTCAAAA
651 TGTATGTGTT CATCCGGGGG ACCTGTAGAC AGAGCACCAG ACTCCTCTAA
701 GGATGGTAGA AACCATCATA AAACAATGGG GCATTATAAC CACAATGACT
751 CTTTCCGGAA AACACAGATA AACTCTTTGA AAAACTCAGT GGCCTGCCCT
801 GGTGCACAAA AGGCTATTAC GTCTTCAGAG GCAGTTGACA GATTTATGCC
851 TAGGACAACA CAACTGCAGG AGCGCAAAAG AAGAAGAGAA GATGATCGTA
901 AACTTGGAACT TTTTCTTCAA ACAAACCCAA CTGGTAATGA GATTATGATT
951 GGACCTCTGT TACCAGACAT CTCTAAAGTG GATATGCACG ATGACTCATT
1001 GAATACAACG GCGAATTTAA TTCGGCATAA ACTTAAAGAG GTATTTTCATC
1051 TGTGCCAAAG CCTCCAGAGG ACAAGCCAGA AGATGTACAT ACAAGTCATC
1101 CATTAAAACA AAGAAGAAGA ATATAGAGTG CCAGCAGCAA CTTAGTATTT
1151 TCTAAAAAGA ACATTTATTA TTTATTTTGA GCCTGTCATT TTAATTCCTC
1201 AAGAGATTTT ACTGCTGGTA TTTTGTGATG CACTCCTCTT TGAATTTTCA
1251 TTCAAGCCAT TTGCTAAAG TCATTCTTTT GTTTTTTGGG AGATGGAGTC
1301 TTGCTCTGTT GCCCAGGCTG GAATGCAGTG GCGTGATCTC GGCTCACTGC
1351 AACCTCCACC TCCCAGGTTT AAGCGATTCT CCTGCCTCAG CCTCCTGAGT
1401 ATCTGGGATT ACAGGCGTGC ACCACCATGC CTGGCTAAGT TTTGTGTTTT
1451 TTTTAGTAGA GATGGGTTT CACCATATTG GTCAGGCTGG TCTCGAACTC
1501 CTGACCTTGT GATACACCTG CCTCAGCCTC CCAAAGGGAT GAGCCACCGC
1551 GCCTGGCCCA TTTCTTCTTT TTTTGACCCA TACTTAATGT TGCAGAAACT
1601 ATTCTGTGCA TAACATTATC TCTCATGTAC AGTAATTATA TGTAAATTAA
1651 TTGAAGCAAA TATGGAAACT TTACAATAGA AATAAAGATA GGCAGCCAGC
1701 GTCTGTTTCC AATTATAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry AC005156 from database EMBL:
Homo sapiens PAC clone DJ1099C19 from 7q21-q22, complete sequence.
Score = 2897, P = 2.4e-154, identities = 583/586
2 exons covering Bp 465-1723

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 24 bp to 1103 bp; peptide length: 360
Category: similarity to unknown protein

```

1 MASSGGEPGS LFDHHVQRAV CDTRAKYREG RRPRAVKVYT INLESQYLLI
51 QGVPAVGVMK ELVERFALYG AIEQYNALDE YPAEDFTEVY LIKFMNLQSA
101 RTAKRKMDEQ SFEGGLLHVC YAPFETVEE TRKKLQMRKA YVVKTTENKD
151 HYVTKKKLVT EHKDTEDFRQ DFHSEMSGFC KAAALNTSAGN SNPYLPYSCE
201 LPLCYFSSKC MCSSGGPVDR APDSSKDGRN HHKTMGHYNH NDSLRLKTQIN
251 SLKNSVACPG AQKAITSSA VDRFMPRTTQ LQERKRRRED DRKLGTFLQT
301 NPTGNEIMIG PLLPDISKVD MHDDSLNTTA NLIRHKLKEV FHLCSLQRT
351 SQKMYIQVIH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23o5, frame 3

TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC F15K20 genomic sequence, complete sequence., N = 2, Score = 114, P = 3.6e-11

>TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC F15K20 genomic sequence, complete sequence.
Length = 227

HSPs:

Score = 114 (17.1 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
Identities = 21/41 (51%), Positives = 29/41 (70%)

Query: 103 AKRKMDEQSFFGGLLHVCYAPFETVEETRKKLQMRKAYVV 143
AKRK+DE SF G L + YAPE+E V +T+ KL+ R+ V+
Sbjct: 51 AKRKLDESSFLGNRLQISYAPEYENVNDTKDKLESRRKEVL 91

Score = 107 (16.1 bits), Expect = 2.6e-10, Sum P(2) = 2.6e-10
Identities = 50/191 (26%), Positives = 83/191 (43%)

Query: 103 AKRKMDEQSFFGGLLHVCYAPFETVEETRKKLQMRKAYVVKTENKDHVYTKKKLVTEH 162
AKRK+DE SF G L + YAPE+E V +T+ KL+ R+ V+ + T + VT+
Sbjct: 51 AKRKLDESSFLGNRLQISYAPEYENVNDTKDKLESRRKEVLARLNPQEKSTSQ--VTKL 108

Query: 163 KDTEDFRQDFHSEMSGFCALNTSAGNSNPYLPYSCELPYFSSKCMCSSGGPVDRAP 222
+ D S + + GN+ P S + YF+S M + V
Sbjct: 109 AGPALTTQTDNVSSQRREMEYQFHR--GNA-PVTRVSSDQE--YFASSSMNQTVKTV---- 159

Query: 223 DSSKDGRRHHKTMGHYNHNDLRLKTQINSLKNSVACPGAQKAITSSAEDRFMPRTTQLQ 282
K + + + +H + + + N + P +Q S R P ++Q+Q
Sbjct: 160 -REKLNKTRREENISLSHCKQIEESG-NQKRLQ---PSSQTQPEESGNQKRLQP-SSQIQ 213

Query: 283 -ERKRRREDDRK 293
+ KR R D+R+
Sbjct: 214 PDLKRTRVDNR 225

Score = 102 (15.3 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
Identities = 22/55 (40%), Positives = 38/55 (69%)

Query: 26 KYREGRRPRAVKVYTINLESQYLLIQGVPAVGVMKELVERFALYGAIQY--NALDE 80
+Y++ P AV+VYT+ ES+Y++++ VPA+G +L+ F YG +E++ LDE
Sbjct: 3 RYKD-ETP-AVRVYTVCDSESRMIVRNVPALGCGDDLMLFMTYGEVEEFAKRLDE 57

Pedant information for DKFZphfbr2_23o5, frame 3

Report for DKFZphfbr2_23o5.3

```

[LENGTH]      360
[MW]           41105.85
[pI]           8.89
[HOMOL]        TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC
F15K20 genomic sequence, complete sequence. 5e-12
[PROSITE]      AMIDATION 1
[PROSITE]      MYRISTYL 2
[PROSITE]      CK2_PHOSPHO_SITE 7

```


[PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 3
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 4.17 %

```

SEQ  MASSSGGEPGSLFDHHVQRAVCDTRAKYREGRRPRAVKVYTINLESQYLLIQGVPAVGVMK
SEG  .....
PRD  cccccccccceeeceeeehhhhhhhhhccccceeeeeeccccceeeeeeccccchhhh

SEQ  ELVERFALYGAIEQYNALDEYPAEDFTEVYLIKFMNLQSARTAKRKMDEQSFFGGLLHVC
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhccccceeeehhhhhhhhhhhhhhhhhccccceeee

SEQ  YAPEFETVEETRKKLQMRKAYVVKTTENKDHVYTKKKLVTEHKDTEDFRQDFHSEMSGFC
SEG  .....
PRD  eccchhhhhhhhhhhhhhhheeeccccceeeeeeccccchhhhhhhhhccccce

SEQ  KAALNTSAGNSNPYLPYSCPLCYFSSKCMCSSGGPVDRAPDSSKDGRNHHKTMGHYNH
SEG  .....
PRD  eeeeecccccccccccccccccecccccccccccccccccccccccccccccccccc

SEQ  NDSLRLKTQINSLKNSVACPGAQKAITSSSEAVDRFMPRTTQLQERKRREDDRKLGTFLOT
SEG  .....
PRD  cccceeeccccccccccccceeeccccceeeccccchhhhhhhhhhhccccceeeec

SEQ  NPTGNEIMIGPLLDPISKVMDHDSLNTTANLIRHKLKEVFHLCQSLQRTSQRMYYIQUIH
SEG  .....
PRD  cccccceeeccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcchhhhhcccc

```

Prosites for DKFZphfbr2_23o5.3

PS00001	185->189	ASN_GLYCOSYLATION	PDOC00001
PS00001	241->245	ASN_GLYCOSYLATION	PDOC00001
PS00001	327->331	ASN_GLYCOSYLATION	PDOC00001
PS00005	99->102	PKC_PHOSPHO_SITE	PDOC00005
PS00005	102->105	PKC_PHOSPHO_SITE	PDOC00005
PS00005	131->134	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	224->227	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00005	251->254	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00006	4->8	CK2_PHOSPHO_SITE	PDOC00006
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	224->228	CK2_PHOSPHO_SITE	PDOC00006
PS00006	266->270	CK2_PHOSPHO_SITE	PDOC00006
PS00006	303->307	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	260->266	MYRISTYL	PDOC00008
PS00009	29->33	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_23o5.3)

DKFZphfbr2_2a2

group: brain derived

DKFZphfbr2_2a2.3 encodes a novel 167 amino acid protein with weak similarity to human 52K autoantigen Ro/SS-A

The novel protein contains a C3HC4 Zinc finger "RING finger" motive. This domain is probably involved in mediating protein-protein interactions. Proteins containing a RING-finger are: mammalian V(D)J recombination activating protein (RAG1), mouse rpt-1, human rfp, human 52 Kd Ro/SS-A protein and others.

No informative BLAST results; no predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to 52K autoantigen Ro/SS-A - human

complete cDNA, complete cds, few EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1376 bp

Poly A stretch at pos. 1355, polyadenylation signal at pos. 1340

```
1 GGGGACTCCA AATTAGAAAG GGGACGTCTA GTGGGTTGCC CGGGAGGGGT
51 GCGGGGAGCG GTCCTGGAAA TAATCTGTCC TCTGTCGCC GGAAGTGGCG
101 AGGTAGTTCC TTCGCGGTGG AGAGACCTGG AATGGCCAAA TATCAAGGTG
151 AAGTTCAAAG TTTGAAACTG GATGATGATT CAGTTATAGA AGGAGTAAGC
201 GACCAAGTAC TTGTGGCAGT TGTGGTCAGT TTCGCTTTGA TTGCTACCCCT
251 GGTATATGCA CTTTTCAGAA ATGTACATCA AAACATTAC CCAGAAAACC
301 AGGAGCTAGT AAGGGTACTT CGAGAACAGC TTCAAACAGA ACAGGATGCA
351 CCTGCTGCCA CTCGACAGCA GTTCTACACT GACATGTA CTCCCATCTG
401 CCTGCACCAA GCCTCCTTCC CGGTGGAGAC CAACTGTGGA CATCTTTTTT
451 GTGGTGCCCTG CATTATTGCT TACTGGCGAT ATGGTTCATG GCTTGGGGCA
501 ATCAGTTGTC CAATCTGTAG ACAAACGGTA ACCTTACTCC TAACAGTATT
551 TGGTGAAGAT GATCAGTCTC AGGATGTTCT GAGATTGCAT CAGGATATTA
601 ATGATTATAA CCGGAGATTC TCAGGGCAAC CCTGATCTAT TATGGAGAGA
651 ATTATGGATC TACCCACTTT ACTGAGGCAT GCATTCAGGG AAATGTTTTC
701 AGTCGGGGGC CTTTCTGGA TGTTCGCAT CAGGATAATA CTTTGTTTAA
751 TGGGAGCTTT TTTCTATCTT ATATCACCTC TAGATTTTGT ACCTGAAGCC
801 TTGTTTGGA TTCTAGGCTT TCTAGATGAT TTCTTTGTCA TCTTTTATT
851 GCTTATCTAC ATCTCTATTA TGTATCGAGA AGTGATAACC CAAAGGCTAA
901 CTAGATGAAA AAGGAAACAA AACTGAGTTT ACTAGGATAT CTGAGCTAAT
951 GTAGAATCAT AACAGAAAGG ACCCATGGCA GTATAAGCA ATGAAGCAAT
1001 GGAGTATTAT CTCACAAATA TAAACCACT ATAAGACAAA CATTGATTA
1051 TCATTTGACA AATACCTAGG TATAACTGGA ATTTTCATGT TTGAAGTTCT
1101 AATATTAAAT TTAGAATTAT AATGATCTAC AGTTGTATCT TGATTCTATG
1151 TTGTCTGGAA AAAAATATGGA ATTATATAAA AAGGGATGCT TTTATATATT
1201 TTTCTTTTCC CCAGAATTAC TTAGATTAAT TAGATGTATA GTAAAAATATT
1251 GTTAAATGTC AGTTTATCCA TCTTATCCTT CTCAGCAGGT ACCTATATGA
1301 TAATATATAG CTGTGAAACT CATCTAAATA TTTTGTGTCC AATAAAATAT
1351 TATATACTAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 132 bp to 632 bp; peptide length: 167
Category: similarity to known protein
Classification: unset

Prosite motifs: ZINC_FINGER_C3HC4 (102-112)

```

1 MAKYQGEVQS LKLDSDSVIE GVSDQVLVAV VVSFALIATL VYALFRNVHQ
51 NIHPENQELV RVLREQLTE QDAPAATROQ FYTDMYCPIC LHOASFPVET
101 NCGHLFCGAC IAYWRYGSW LGAISCPICR QTVTLTLTVF GEDDQSQDVL
151 RLHQDINDYN RRFSGQP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2a2, frame 3

TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A, N = 1, Score = 194, P = 2e-15

PIR:T05222 hypothetical protein F17I5.130 - Arabidopsis thaliana, N = 1, Score = 159, P = 1.4e-10

TREMBLNEW:AB025011_1 gene: "TRIF"; product: "Trif-d"; Mus musculus mRNA for Trif-d, complete cds., N = 1, Score = 108, P = 2.6e-06

PIR:A37241 52K autoantigen Ro/SS-A - human, N = 1, Score = 115, P = 5e-05

>TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A
Length = 283

HSPs:

Score = 194 (29.1 bits), Expect = 2.0e-15, P = 2.0e-15
Identities = 52/149 (34%), Positives = 78/149 (52%)

```

Query:   16 DSVIEGVSDQVLVAVVVSFALIATLVYALFRNVHQNIHPENQELVRVLREQLTEQDAPA 75
          D +E ++ Q+ +A+ V F ++ + A Q E R Q+ T++
Sbjct:   41 DPVE-LATQITMAIVIF-IVKAIFDAWQSRRRQRAASRMDENAE--RNQIITQRRISE 96

Query:   76 ATRQQFYTDMYCPICLHQASFPVETNCGHLFCGACIIAYWRYGSWLGA-ISCPICRQTVT 134
          A Q + CPICL ASFPV T+CGH+FC CII YW+ + C +CR T
Sbjct:   97 ALHQSSHE---CPICLANASFPVLTDCGHIFCCECIIQYWQQSKAIVTPCDCAMCRSTFY 153

Query:   135 LLLTV----FGEDDQSQDVLRLHQ-DINDYNRRFS 164
          +LL V G +++ D ++ + I+DYNRRFS
Sbjct:   154 MLLPVHWPTMGTSEETDDHIQENNIRIDDYNRRFS 188

```

Pedant information for DKFZphfbr2_2a2, frame 3

Report for DKFZphfbr2_2a2.3

```

[LENGTH]      167
[MW]           18941.65
[pI]           4.91
[HOMOL]        TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A 1e-13

[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]       30.19 peroxisomal organization [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YLR323c] 2e-04
[BLOCKS]       BL00518 Zinc finger, C3HC4 type, proteins
[PROSITE]      ZINC_FINGER_C3HC4 1
[PFAM]         Zinc finger, C3HC4 type (RING finger)
[KW]           Irregular
[KW]           3D
[KW]           LOW_COMPLEXITY 6.59 %

```

```

SEQ      MAKYQGEVQSLKLDSDSVIEGVSDQVLVAVVVSFALIATLVYALFRNVHQNIHPENQELV
SEG      .....XXXXXXXXXXXXX.....
lrmd-    .....

SEQ      RVLREQLTEQDAPAATROQFYTDMYCPICLHQASFPVETNCGHLFCGACIIAYWRYGSW
SEG      .....
lrmd-    .....HHHHHHBTTTTEETTTEETTEEEHHEHHHH---HHHHH

SEQ      LGAISCPICRQTVTLTLTVFGEDDQSQDVLRLHQDINDYNRRFSGQP

```

SEG
1rmd- HCCB-TTTT.....

Prosite for DKFZphfbr2_2a2.3

PS00518 102->112 ZINC_FINGER_C3HC4 PDOC00449

Pfam for DKFZphfbr2_2a2.3

HMM_NAME Zinc finger, C3HC4 type (RING finger)

HMM *CPICFcTFQlDyPWPfdePmMlPCgHsFCypCIrrW.....CP
CPIC L+ P++++CGH+FC +CI+ + CP

Query 87 CPIC-----LHQ---ASFPVETNCGHLFCGACIIAYWRYGSWLGAISCP 127

HMM mC*

Query 128 IC 129

DKF2phfbr2_2b17

group: transmembrane protein

DKF2phfbr2_2b17 encodes a novel 285 amino acid protein with similarity to D. melanogaster 30K protein.

The protein contains 3 transmembrane regions.
No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to Drosophila hypothetical 30K protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 3

Sequenced by Qiagen

Locus: unknown

Insert length: 1426 bp

Poly A stretch at pos. 1345, polyadenylation signal at pos. 1330

```

1 GGGGGTATTT CCAAGGACTC CAAAGCGAGG CCGGGGACTG AAGGTGTGGG
51 TGTCGAGCCC TCTGGCAGAG GGTAACTCTG GGTCAAATGC ACGGATTCTC
101 ACCTCGTACA GTTACGCTCT CCCGCGGCAC GTCCGCGAGG ACTTGAAAGTC
151 CTGAGCGCTC AAGTTTGTCC GTAGGTCGAG AGAAGGCCAT GGAGGTGCCG
201 CCACCGGCAC CGCGGAGCTT TCTCTGTAGA GCATTGTGCC TATTTCCCCG
251 AGTCTTTGCT GCCGAAGCTG TGACTGCCGA TTCGGAAAGTC CTTGAGGAGC
301 GTCAGAAAGCG GCTTCCCTAC GTCCCAGAGC CCTATTACCC GGAATCTGGA
351 TGGGACCGCC TCCGGGAGCT GTTTGGCAA GATGAACAGC AGAGAATTTT
401 AAAGGACCTT GCTAATATCT GTAAGACGGC GGCTACAGCA GGCATCATTG
451 GCTGGGTGTA TGGGGGAATA CCAGCTTTTA TTCATGCTAA ACAACAATAC
501 ATTGAGCAGA GCCAGGCAGA AATTTATCAT AACCGGTTTG ATGCTGTGCA
551 ATCTGCACAT CGTGTGCCA CACGAGGCTT CATTCGTTAT GGCTGGCGCT
601 GGGGTTGGAG AACTGCAGTG TTTGTGACTA TATTCAACAC AGTGAACACT
651 AGTCTGAATG TATACCGAAA TAAAGATGCC TTAAGCCATT TTGTAATTGC
701 AGGAGCTGTC ACGGGAAGTC TTTTATGAT AAACGTAGGC CTGCGTGGCC
751 TGGTGGCTGG TGGCATAATT GGAGCCTTGC TGGGCACTCC TGTAGGAGGC
801 CTGCTGATGG CATTTAGAAA GTACTCTGTG GAGACTGTTC AGGAAAGAAA
851 ACAGAAAGAT CGAAAGGCAC TCCATGAGCT AAAACTGGAA GAGTGGAAAG
901 GCAGACTACA AGTTACTGAG CACCTCCCTG AGAAAATTGA AAGTAGTTTA
951 CAGGAAGATG AACCTGAGAA TGATGCTAAG AAAATTGAAG CACTGCTAAA
1001 CCTTCCTAGA AACCTTTCAG TAATAGATAA ACAAGACAAG GACTGAAAAGT
1051 GCTCTGAATG TGAACCTCAC TGGAGAGCTG AAGGGAGCTG CCATGTCCGA
1101 TGAATGCCAA CAGACAGGCC ACTCTTTGGT CAGCCTGCTG ACAAATTTAA
1151 GTGCTGGTAC CTGTGGTGGC AGTGGCTTGC TCTTGCTCTT TTCTTTTCTT
1201 TTTAACTAAG AATGGGGCTG TTGTACTCTC ACTTTACTTA TCCTTAAATT
1251 TAAATACATA CTTATGTTTG TATTAATCTA TCAATATATG CATACATGAA
1301 TATATCCACC CACCTAGATT TTAAGCAGTA AATAAACAT TTGCAAAAAG
1351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HSG19630 from database EMBL:

human STS A001T27.

Score = 961, P = 1.2e-36, identities = 193/194

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 189 bp to 1043 bp; peptide length: 285
Category: similarity to unknown protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2 2b17, frame 3

>PIR:JQ1024 hypothetical 30K protein (DmRP140 5' region) - fruit fly
(*Drosophila melanogaster*)
Length = 261

HSPs :

```

Query:      30 ADSEVLEERQKRLPYVPEPPYPESGWDRRLRELFGKDEQQRISKDLANICKTAAATAGIIGW 89
            AD V +E + ++ E+G +RL+++F DE I +L ++ + +IG
Sbjct:      23 ADEIVDKENKTYKAFASKPPEETGLERLKLQMFITDEFSGISFELNSVYQAGFLGFLIGA 82

Query:      90 VYGGIPAFIHAQQQYIEQSAQYIYHNRFDVQSAHRAATRGFIYRWGWRWGTAVFVTF 149
            +YGG+ A ++E+QA ++ + FDA + T F + G++WGWR +F T +
Sbjct:      83 IYGGVTQSRVAYMNFEMENNQATAFKSHFDAKKKLQDQFTVNFAGKGGFKWGRVGLFTTSY 142

Query:      150 NTVNTSLNVRNKDALSHFVIAGAVTGSLFRINVLRLGLVAGGIIGALLGTPVGGLLMAF 209
            + T ++VYR K ++. ++ AG++TGS L++++GLRG+ AGGIIG LG G +
Sbjct:      143 FGIITCMSVYRGKSSIIYEYLAAGSITGSLYKVS LGLRGMAAGGIIGGLGGVAGVTSLLL 202

Query:      210 QKYSGETVQERKQKDRKALHELKLEEWKGR LQVTEHLPEKIESSLQDEPE 260
            K SG +++E ++ ++K RL E++ + + +++ PE
Sbjct:      203 MKASGTSMEE-----VRYWQYKRLDRDENIOQAFKKLTEDENPE 242

```

Pedant information for DKF2phfbr2 2b17. frame 3

Report for DKFZphfbr2 2b17.3

```
[LENGTH]          285
[MW]               32177.88
[pI]               8.65
[HOMOL]            PIR:JQ1024 hypothetical 30K protein (DmRPl40 5' region) - fruit fly (Drosophila
melanogaster) 7e-20
[PROSITE]          MYRISTYL          7
[PROSITE]          CK2_PHOSPHO_SITE      5
[PROSITE]          ASN_GLYCOSYLATION     1
[KW]               SIGNAL PEPTIDE 25
[KW]               TRANSMEMBRANE 3
[KW]               LOW COMPLEXITY      5.96 %
```

199

```

SEG      ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      eeeeeeeeeeeeeeeeeeeeecccccccccccccccccccccccccccccccccccccccc
MEM      mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

```

```

SEQ      QVTEHLPEKIESSLQEDEPENDAKKIEALLNLPRNPSVIDKQDKD
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

```

Prosites for DKFZphfbr2_2b17.3

PS00001	153->157	ASN_GLYCOSYLATION	PDOC00001
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	172->178	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2b17.3)

DKFZphfbr2_2b5

group: cell structure and motility

DKFZphfbr2_2b5 encodes a novel 957 amino acid protein with strong similarity to collagens.

The novel protein contains the typical (xxG)n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain.

The new protein can find application in modulation of connective tissue, bone and cartilage development and maintenance.

similarity to collagen proteins

shows typical (xxG)n repeat of collagen proteins
[PFAM] von Willebrand factor type A domain

Sequenced by Qiagen

Locus: /map="6"

Insert length: 4160 bp

Poly A stretch at pos. 4141, polyadenylation signal at pos. 4119

```
1 GGGGGCCCCG TGCAGGGAGA ACGGACTCCG GCGGGAGGGC AGCCAATCCG
51 TTTCAGCGCA GGTCTTGCTC GGGTTGGGCT TGCCACTGCC TGGAAACATAC
101 CTGTCCCCCT GCGCGAACAC TCAGCTGGCT GCGACCGCAA CCCCAGCCT
151 GGACACTGCG CCAGGAATCC TAAAACCAA ATATTAGAAC GAAAACAGAA
201 ACATGGCTCA CTATATTACA TTCTCTGCA TGGTTTGGT GCTGCTTCTT
251 CAGAATTCGT TGTAGCTGA AGATGGGGAA GTAAGATCAA GTTGTCGTAC
301 TGCTCCGACA GATTTAGTTT TCATCTTAGA TGGCTCTTAT AGTGTTGGCC
351 CAGAAATCTT TGAATAGTG AAAAAGTGGC TTGTCAATAT CACAAAAAAC
401 TTTGACATAG GGCCGAAGTT TATTCAGTT GGAGTGGTTC AATATAGTGA
451 CTACCTCTGT CTGGAGATTC CTCTCGGAAG CTATGATTCA GGAGAACATT
501 TGACGGCAGC AGTGGAAATCC ATACTCTACT TAGGAGGAAA CACAAAGACA
551 GGGGAAGGCCA TCCAGTTTGC GCTCGATTAC CTTTTTGACA AGTCCTCAGC
601 ATTTCTGACT AAGATAGCAG TGGTACTTAC GGATGGCAAG TCCCAAGATG
651 ACGTCAAGGA TGCAGCTCAA GCAGCAAGAG ATAGTAAGAT AACATTATTT
701 GCTATTGGTG TTGGTTCAGA AACAGAAGAT GCCGAACCTA GAGCTATTGC
751 CAACAAGCCT TCGTCTACTT ATGTGTTTTA TGTGGAAGAC TATATTGCAA
801 TATCCAAATC AAGGGAAGTG ATGAAGCAGA AACTTTGTGA AGAATCTGTC
851 TGTCCAACAC GAATTCAGT GGCAGCTCGT GATGAAAGGG GATTTGATAT
901 TCTTTTGGGT TTAGATGTAA ATAAAAAGGT TAAGAAAAGA ATACAGCTTT
951 CACCAAAAAA GATAAAAGGA TATGAAGTAA CATCAAAAGT TGATTATATCA
1001 GAACTCACAA GCAATGTTTT CCCAGAAGGT CTCCTCCAT CATATGTATT
1051 TGTGTCTACT CAAAGATTTA AAGTCAAGAA AATTTGGGAT TTATGGAGAA
1101 TATTAACTAT TGATGGAAGG CCACAAATAG CAGTTACCTT AAATGGTGTG
1151 GACAAAATCT TATTATTTAC AACAAACAGC GTAATTAATG GCTCACAAGT
1201 GGTACCTTTT GCTAACCTC AAGTTAAGAC GTTGTTTGAT GAAGGCTGGC
1251 ACCAAATTCG TCTCTTAGTA ACAGAACAAG ATGTGACTTT GTATATTGAT
1301 GACCAACAAA TTGAAAACAA GCCCTTACAT CCAAGTTTAG GGATCTTGAT
1351 CAATGGGCAA ACCCAAATTG GAAAATATTC TGGAAAAGAA GAAACTGTTC
1401 AGTTTGATGT CCAAAAGTTG CGAATCTACT GTGACCCAGA ACAGAACAAC
1451 CGGGAGACAG CATGTGAGAT TCCTGGATT TATGGAGAGT GCCTTAATGG
1501 TCCCAAGTAT GTAGGTCAA CTCCAGCTCC CTGTATTTGT CCTCCGGGAA
1551 AACCAGGACT TCAAGGCCCC AAAGGTGACC CTGGACTGCC TGGGAACCTT
1601 GGCTACCTGT GACAACCTGG TCAAGATGGT AAGCCTGGAT ATCAGGGAAT
1651 TGCAAGGAGA CCAGGTGTTT CAGGATCTCC AGGAATACAA GGAGCTCGAG
1701 GACTACCAAG TTACAAAGGA GAACCAGGGC GAGATGGTGA CAAGGGTGAT
1751 CGTGGACTTC CTGGTTTTCC TGGGCTTCAT GGCATGCCAG GATCAAAGGG
1801 TGAAATGGGT GCCAAGGAG ACAAAGGATC ACCTGGATT TATGGCAAAA
1851 AGGGTGCAAA AGGTGAAAAG GGAATGCTG GCTTCCTGG CCTCCCTGGA
1901 CCTGTGGAG AACCAAGGAG ACATGGAAG GATGGATTAA TGGGTAGTCC
1951 CGGTTTCRAA GGAGAGCAG GATCCCTGG TGCTCCGGGG CAGGATGGAA
2001 AAGCAACCGG GCCTGGAATC CCAGGATTTT CTGGAACCGG AGGATTAAATG
2051 GGGCAAAAGG GAGAAATTGG GCCTCCAGGA CAGCAAGGAA AAAAAGGAGC
2101 CCCAGGGATG CTGTTTAA TGGGAAGCAA TGGCTCACC GGCCAGCCTG
2151 GAACACCGGG ATCTAAGGGA AGCAAAGGTG AACCTGGAAT TCAAGGGATG
2201 CCTGGGGCTT CAGGGCTCAA GGGAGAACCA GGAGCAACGG GTTCCCAGG
2251 AGAACCAGGA TACATGGGTT TACCCGGGAT TCAAGGAAAA AAGGGGGACA
2301 AAGGAAATCA AGGTGAAAAA GGTATTCAAG GTCAAAAGGG AGAAAAATGGA
2351 AGACAGGGAA TTCCAGGGCA ACAGGGAATT CAAGGCCATC ATGGTGCAAA
2401 AGGAGAGAGA GGTGAAAAGG GAGAACCTGG TGTCCGAGGT GCCATTGGAT
2451 CAAAAGGAGA ATCTGGGGTG GATGGCTTGA TGGGGCCCCG AGGTCCTAAG
2501 GGGCAACCTG GGGATCCAGG TCCTCAGGGA CCCCAGGTT TGGATGGGAA
2551 GCCCGGAAGA GAGTTTTCAG AACATTTTAT TCGACAAGTT TGCACAGATG
```



```

2601 TAATAAGAGC CCAGCTACCA GTCTTACTTC AGAGTGGGAG AATTAGAAAT
2651 TGTGATCATT GCCTGTCCCA ACATGGCTCC CCGGGTATTC CTGGGCCACC
2701 TGGTCCGATA GGCCAGAGG GTCCAGAGG ATTACCTGGT TTGCCAGGAA
2751 GAGATGGTGT TCCTGGATTA GTGGGTGTCC CTGGACGTCC AGGTGTCAGA
2801 GGATTAAAG GCCTACCAGG AAGAAATGGG GAAAAAGGGA GCCAAGGGTT
2851 TGGGTATCCT GGAGAACAAAG GTCCTCCTGG TCCCCCAGGT CCAGAGGGCC
2901 CTCCTGGAAT AAGCAAAGAA GGTCTCCAG GAGACCCAGG TCTCCTGGC
2951 AAAGATGGAG ACCATGGAAA ACCTGGAATC CAAGGGCAAC CAGGCCCCCC
3001 AGGCATCTGC GACCCATCAC TATGTTTATG TGTAATTGCC AGAAGAGATC
3051 CGTTCAGAAA AGGACCAAAC TATTAGTGTC TGATGCCTCA TTCAGCAGCC
3101 TAGGCATGGT GCTTTTCTG TGGTCTTTG CATCTCAGGA AGATAACCAA
3151 CAGTATCCCT TGAAGAGAAA CTTAAGTACC TCGGTGTTT TATTTTTTTT
3201 TTCTTATGGA AAAAAATATA AAAGATCACA TATACTGATT TTAAGGCTC
3251 CTCAGTCATT TGGAGCCCTT GGATTAGCAG CATTAATTA ATCTCAAGGG
3301 TTTCTTGTA AGTCCATTTA TGTTAATCAA AGTTGAATAT AAAAATCCAC
3351 CATGTCCTGT TAGCCAGTCA GTTTTAGTCA CTGTGAAATA TTTACATTTC
3401 AGCCTCCATG CAGTAGAGAT TTGAGTTTAA TTTATGTCC ATGTGACTTT
3451 CATGTTTCTT ATCTCATAGC TCATGCTACT ACATAAGCCA AAACATGTAT
3501 CTCATCATTG GAAGTAAGAT CAGGGCTGAT ATTCACCTGG GATAGACAGT
3551 ATTGGTGAAC TACTCATTTA CTACAGTGTC TCAGCCTTGA TAAAGGGCAG
3601 TGGATTGCCT GTTGTTCGGT GTTGTGAATA GCACCTCTGA ATAAGATTAG
3651 AGTGTTCCTT AATTCATTTC AAACCTCTAA ATTAGATTAA TGGTGGTGCT
3701 AAGAAAGAGT ATTAATTACT TTGGGAATGG TCAAAATTA CATTAATAAC
3751 ATTTAGACA AAAAGTTTCA TTGTACATTC AAAGAAATG TAAGTTTGA
3801 AGTACTAAAA GACTATTTTA TACTTGTGA TTAATCGGAA TGTTTGTGT
3851 ATGCCTTCAT TTTCCATTTC ACTTATATGT GCATGTCCAT ATATGTTAAT
3901 TTTCAATTGA GCAAAGCTAA TGGAAATAAA GCTAATGCTC TAGTTGAAAG
3951 AAAAGGAAAA CTCCTGAAAT CCTAGAATGT CTTGTTATTT TTAGCTGACT
4001 GTAAATATT ATGAACAGTC TTTGTGTATT GTGCTTAATG CTTTGTAAAG
4051 AAACAGAAAT TGAATATTT CATCCTTGTC ATGCTCAAAA TTTTGTTACA
4101 TGCTTGTAT TCAGAGTATA ATAAAGTTT GTACAGGCCT GAAAAAATA
4151 AAAAAAATA

```

BLAST Results

Entry HS682J15 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 682J15
 Score = 6240, P = 0.0e+00, identities = 1256/1263
 13 exons matching Bp 2015-4118

Entry HS708F5 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 708F5
 Score = 2775, P = 1.0e-221, identities = 739/912
 10 exons matching Bp 5-1745

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 203 bp to 3073 bp; peptide length: 957
 Category: similarity to known protein

```

1 MAHYITFLCM VLVLLQNSV LAEDGEVRSS CRTAPTDLVF ILDGYSVGP
51 ENFEIVKKWL VNITKNFDIG PKFIQVGVVQ YSDYPVLEIP LGSYDSGEHL
101 TAAVESILYL GGNTKTGKAI QFALDYLFDK SSRFLTIAV VLTGKSQDD
151 VKDAAQAARD SKITLFAIGV GSETEDAELR AIANKPSSSTY VFYVEDYIAI
201 SKIREVMKQK LCEESVCPTR IPVAARDERG FDILLGLDVN KVKKKRIQLS
251 PKKIKGYEVT SKVDLSELT NVFPEGLPPS YVFVSTQRFK VKKIWDLWRI
301 LTIDGRPQIA VTLNGVDKIL LEFTTSVING SQVVFANPQ VKTLFDEGWH
351 QIRLLVTEQD VTLYIDDQOI ENKPLHPVLG ILLNGQTQIG KYSKEETVQ
401 FDVQKLRIYC DPEQNNRETA CEIPGFNGEC LNGPSDVGST PAPCICPPGK
451 PGLQGPKGDP GLPKNPGYPG QPGQDGKPGY QGIAGTPGVP GSPGIQGARG
501 LPGAQKGEPR DDKGDRGLP GFPLHGMMPG SKGEMGAKGD KSGPGFYGKK
551 GAKGEKGNAG FPGLPGPAGE PGRHGKDGML GSPGFKGEAG SPGAPGQDGT
601 RGEPIPGFP GNRGLMGQKG EIGPPGQGGK KGAPGMPGLM GSNGPSGQPG
651 TPQSGKSGKE PGIQGMPCAS GLKGEFGATG SPGEFGYMG LPIQGGKGDG
701 GNQGEKGIQK QKGENGRQGI PQQGIQGHG GAKGERGEKG EPGVRGAIGS
751 KGESGVVDGLM GPAGPKGQPG DPGPQGPGL DGKPGREFSE QFIRQVCTDV
801 IRAQLPVLLQ SGRIRNCDHC LSQHGSPGIP GPPPGIGPEG PRGLPGLPGR

```

851 DGVPGVLGVGP GRPGVRLKKG LPGRNGEKGS QGFGYPGEQG PPGPPGPEGP
 901 PGISKEGPPG DPGLPGKDGD HGKPGIQGQP GPPGICDPSL CFSVIARRDP
 951 FRKGPNY

BLASTP hits

Entry HSCOL7A1X_1 from database TREMBL:
 gene: "COL7A1"; product: "collagen type VII"; Homo sapiens (clones:
 CW52-2, CW27-6, CW15-2, CW26-5, 11-67) collagen type VII intergenic
 region and (COL7A1) gene, complete cds.
 Score = 949, P = 3.4e-122, identities = 237/553, positives = 281/553

Entry CAL7 HUMAN from database SWISSPROT:
 COLLAGEN ALPHA 1(VII) CHAIN PRECURSOR (LONG-CHAIN COLLAGEN) (LC
 COLLAGEN). >TREMBL:HSCOL7A1_1 gene: "COL7A1"; product: "alpha-1 type
 VII collagen"; Human alpha-1 type VII collagen (COL7A1) mRNA, complete
 cds.
 Score = 949, P = 3.6e-122, identities = 237/553, positives = 281/553

Alert BLASTP hits for DKFZphfbr2_2b5, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2b5, frame 2

Report for DKFZphfbr2_2b5.2

[LENGTH] 957
 [MW] 99413.38
 [pI] 8.49
 [HOMOL] PIR:A40020 collagen alpha 1(XII) chain precursor - chicken 9e-90
 [BLOCKS] BL01119B Copper-fist domain proteins
 [BLOCKS] BL00313B
 [BLOCKS] BL01113A Clq domain proteins
 [BLOCKS] BL00420A Speract receptor repeat proteins domain proteins
 [SCOP] d1zoob 3.45.1.1.1 Integrin CD11a/CD18 (LFA-1) [Human (Hom 2e-58
 [SCOP] dlido 3.45.1.1.2 Integrin CR3 (CD11b/CD18), alpha subunit [Huma 8e-62
 [EC] 3.1.1.7 Acetylcholinesterase 7e-24
 [PIRKW] blocked amino end 1e-43
 [PIRKW] duplication 7e-46
 [PIRKW] cornea 1e-35
 [PIRKW] lung 2e-40
 [PIRKW] leukocyte 1e-42
 [PIRKW] skin 1e-40
 [PIRKW] transmembrane protein 1e-37
 [PIRKW] cartilage 3e-59
 [PIRKW] hydroxylysine 4e-62
 [PIRKW] connective tissue 3e-43
 [PIRKW] triple helix 5e-82
 [PIRKW] homotrimer 2e-37
 [PIRKW] bone 6e-40
 [PIRKW] Alport syndrome 1e-42
 [PIRKW] laminin binding 2e-40
 [PIRKW] liver 2e-40
 [PIRKW] glycoprotein 5e-82
 [PIRKW] carboxylic ester hydrolase 7e-24
 [PIRKW] disulfide bond 7e-46
 [PIRKW] cell binding 7e-46
 [PIRKW] heterotrimer 4e-62
 [PIRKW] calcium binding 8e-28
 [PIRKW] alternative splicing 5e-82
 [PIRKW] coiled coil 5e-82
 [PIRKW] basement membrane 7e-46
 [PIRKW] trimer 5e-82
 [PIRKW] pyroglutamic acid 3e-43
 [PIRKW] hydroxyproline 4e-62
 [PIRKW] extracellular matrix 5e-82
 [PIRKW] chondroitin sulfate proteoglycan 6e-41
 [PIRKW] sulfoprotein 7e-39
 [PIRKW] kidney 1e-42
 [PIRKW] angiogenesis inhibitor 6e-36
 [PIRKW] Ehlers-Danlos syndrome 2e-40
 [SUPFAM] fibronectin type III repeat homology 5e-82
 [SUPFAM] scavenger receptor cysteine-rich domain homology 1e-37
 [SUPFAM] C-type lectin homology 6e-30
 [SUPFAM] collagen alpha 2(I) chain 5e-40
 [SUPFAM] collagen alpha 1(I) chain 6e-44

[SUPFAM] fibrillar collagen carboxyl-terminal homology 6e-44
 [SUPFAM] animal Kunitz-type proteinase inhibitor homology 2e-38
 [SUPFAM] fibronectin type II repeat homology 6e-21
 [SUPFAM] complement C1q carboxyl-terminal homology 1e-38
 [SUPFAM] collagen alpha 3(VI) chain 2e-31
 [SUPFAM] collagen alpha 1(IV) chain 7e-46
 [SUPFAM] collagen alpha 1(VI) chain 2e-37
 [SUPFAM] von Willebrand factor type C repeat homology 6e-44
 [SUPFAM] unassigned collagens 4e-62
 [SUPFAM] von Willebrand factor type A repeat homology 5e-82
 [SUPFAM] collagen alpha 1(XIV) chain 5e-82
 [SUPFAM] pulmonary surfactant protein D 6e-30
 [SUPFAM] collagen alpha 1(V) chain 7e-39
 [SUPFAM] collagen alpha 1(VIII) chain 1e-38
 [SUPFAM] EGF homology 1e-35
 [PROSITE] AMIDATION 3
 [PROSITE] MYRISTYL 14
 [PROSITE] CK2_PHOSPHO_SITE 13
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 2
 [PFAM] von Willebrand factor type A domain
 [KW] Irregular
 [KW] 3D
 [KW] SIGNAL_PEPTIDE 23
 [KW] LOW_COMPLEXITY 24.24 %

SEQ MAHYITFLCMVLVLLQLNSVLAEDGEVRSSCRTAPTDLVFILDGSSVGPENFEIVKKWL
 SEG
 latzBCCCEEEEEEECCCCCHHHHHHHHHHH

SEQ VNITKNFDIGPKFIQGVVQYSDYPVLEIPLGSYDSGEHLTAAVESILYLGGNKTGKAI
 SEG
 latzB HHHHHHCCBTTTTEEEEEETTTTETTTTTHHHHHHHHHHCCCCCCCCCHHHHH

SEQ QFALDYLFDKSSRFLTKIAVVLTDGKSQDDVKDAAQAARDKITLFAIGVGSETEDAELR
 SEG
 latzB HHHHHHHHCCCTTTTTEEEEEEECCCTTTTHHHHHHHHHHHCEEEEEEECCCCCHHHHH

SEQ AIANKPSSTYVFYVEDYIAISKIREVMKQKLCESVCPTRIPVAARDERGFIDILLGLDVN
 SEG
 latzB HHHGGGGGGGCECCHHHHHHHHHCHHHHHHHHH.....

SEQ KKVKKRIQLSPKKIKGYEVTSKVDLSELTSNVPEGLPPSYVVFSTQRFKVKKIWDLWRI
 SEG
 latzB

SEQ LTIDGRPQIAVTLNGVDKILLFTTTSVINGSQVVTFANPQVKTLFDEGWHQIRLLVTEQD
 SEG
 latzB

SEQ VTLYIDDQQIENKPLHPVLGILINGQTQIGKYSKKEETVQFDVQKLRIYCDPEQNNRETA
 SEG
 latzB

SEQ CEIPGFNGECLNGPSDVGSTPAPCICPPGKPGLGKDPGLPGNPGYPGPGQDQKPGY
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 latzB

SEQ QGIAGTPGVPGSPGIQARGPLPGYKGEPRDGDGDRGLPGFPGLHGMPSKSGEMGAKGD
 SEG XX.....
 latzB

SEQ KGSPGFYGGKGAKEGKNAGFPGLPGPAGEPRHGKDGLMSPGFKGEAGSPGAPGQDGT
 SEGXXXXXXXXXXXXXXXX.....
 latzB

SEQ RGEPIPGFPGNRLMGQKGEIGPPGQQGKKGAPGMPGLMGSNGSPGQPGTSGSGSGE
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 latzB

SEQ PGIQMPGASGLKGEPGATGSPGEPGYMGLPGIQKKGDKGNQGEKGIQGGKGENGRQGI
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 latzB

SEQ PGQQGIQHHGAKGERGEKGEVGRGAIGSKGESVDGLMGPAGPKGQPGDPPGPPGL
 SEG XXXXXXXXXXXX.....XXXXXXXXXXXXXXXXXXXX
 latzB

SEQ DGKPGREFSEQFIRQVCTDVIRAQLPVLLQSGRIRNCDHCLSQHSGSPGIPGPPGPIGPEG
 SEG XXXXX.....XXXXXXXXXXXXXXXXXXXX

```

latzB .....
SEQ   PRGLPGLPGRDGVPLVGVPGRPGVRGLKGLPGRNGEKGSQGFYPGEQGPPGPPGEGP
SEG   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
latzB .....
SEQ   PGISKEGPPGDPLPGKGDGDKGKPGIQGQPGPPGICDPSLCFSVIARRDPFRKGPNY
SEG   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
latzB .....

```

Prosite for DKFZphfbr2_2b5.2

```

PS00001      62->66   ASN_GLYCOSYLATION      PDOC00001
PS00001      329->333 ASN_GLYCOSYLATION      PDOC00001
PS00005       30->33   PKC_PHOSPHO_SITE      PDOC00005
PS00005      116->119 PKC_PHOSPHO_SITE      PDOC00005
PS00005      131->134 PKC_PHOSPHO_SITE      PDOC00005
PS00005      250->253 PKC_PHOSPHO_SITE      PDOC00005
PS00005      260->263 PKC_PHOSPHO_SITE      PDOC00005
PS00005      286->289 PKC_PHOSPHO_SITE      PDOC00005
PS00005      393->396 PKC_PHOSPHO_SITE      PDOC00005
PS00005      811->814 PKC_PHOSPHO_SITE      PDOC00005
PS00006      147->151 CK2_PHOSPHO_SITE      PDOC00006
PS00006      172->176 CK2_PHOSPHO_SITE      PDOC00006
PS00006      261->265 CK2_PHOSPHO_SITE      PDOC00006
PS00006      343->347 CK2_PHOSPHO_SITE      PDOC00006
PS00006      357->361 CK2_PHOSPHO_SITE      PDOC00006
PS00006      393->397 CK2_PHOSPHO_SITE      PDOC00006
PS00006      419->423 CK2_PHOSPHO_SITE      PDOC00006
PS00006      531->535 CK2_PHOSPHO_SITE      PDOC00006
PS00006      600->604 CK2_PHOSPHO_SITE      PDOC00006
PS00006      657->661 CK2_PHOSPHO_SITE      PDOC00006
PS00006      681->685 CK2_PHOSPHO_SITE      PDOC00006
PS00006      750->754 CK2_PHOSPHO_SITE      PDOC00006
PS00006      754->758 CK2_PHOSPHO_SITE      PDOC00006
PS00008       92->98   MYRISTYL          PDOC00008
PS00008      112->118 MYRISTYL          PDOC00008
PS00008      236->242 MYRISTYL          PDOC00008
PS00008      276->282 MYRISTYL          PDOC00008
PS00008      380->386 MYRISTYL          PDOC00008
PS00008      494->500 MYRISTYL          PDOC00008
PS00008      527->533 MYRISTYL          PDOC00008
PS00008      596->602 MYRISTYL          PDOC00008
PS00008      638->644 MYRISTYL          PDOC00008
PS00008      650->656 MYRISTYL          PDOC00008
PS00008      653->659 MYRISTYL          PDOC00008
PS00008      665->671 MYRISTYL          PDOC00008
PS00008      743->749 MYRISTYL          PDOC00008
PS00008      746->752 MYRISTYL          PDOC00008
PS00009      547->551 AMIDATION          PDOC00009
PS00009      628->632 AMIDATION          PDOC00009
PS00009      694->698 AMIDATION          PDOC00009

```

Pfam for DKFZphfbr2_2b5.2

```

HMM_NAME      von Willebrand factor type A domain
HMM            *DIVFLIDGSdSIGpqNFNrmKDFierMMERMDIgPDwIRVGVVQYSdNP
               D+VF++DGS S+GP NF+++K+ ++++ ++DIGP+ I+VGVVQYSD P
Query          37  DLVFILDGSYSVGPENFEIVKKWLVNITKNFDIGPKFIQGVVQYSDYP      85
HMM            RqEmrFmFNDYQNKeEILQaIqqMMYWMgggTNTGeAIQYVvrNMFwee
               E +++ Y + E++++A+ ++ ++GG T+TG AIQ+++++F +++
Query          86  VLE--IPLGSYDSGEHLTAIVESIL-YLGGNTKTGKAIQFALDYLFDKSS    132
HMM            GmRWenvPQVMIIITDGRSQDDIRDPIneMrrmaGIqvFaIGIGNhDNNn
               + ++++++TDG+SQDD++D+++++R+ I+ FAIG+G
Query          133 RF----LTKIAVVLTDGKSQDDVKDAAQAARD-SKITLFAIGVGSETE--    175
HMM            WeELReIASePdEdHVfYvDfFeeLdnMqeql*
               +ELR IA++P++ +VFYV+D+ +++ ++E +
Query          176 DAELRAIANKPSSTYVFYVEDYIAISKIREVM      207

```

DKF2phfbr2_2c1

group: brain derived

DKF2phfbr2_2c1 encodes a novel 697 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 3973 bp

Poly A stretch at pos. 3914, polyadenylation signal at pos. 3900

```
1  GGGGGGATT CGGCGGCGGA AACATGGCGG TCGCGGCCGG GCCGGTAACG
51  GAGAAAGTTT ACGCCGACAC TGGCCTGTAT TAGCGCGTAT GGCCTCGGGC
101 CCTCGTTCCC CAAGGCGTGC CGCCTCCCTG TTCTCAGTCG CAGGCTGAAG
151 CCTGTCTGCG TCTCCTCCTT TTTGGTTTGG TTTTGGAACT GACTCCGAGG
201 GTTGGGAGAG CGCGTTGGTG GCGACGGCCG AGTCAGATCA CTATAAACAA
251 AATTTCACCA AGAGAAAATG TTGAAATAGG AGTTGCGGAT ACATTGGATA
301 TACTGGATGA AATACAAGCG GTTAATTTT GTAACGTGAG GGAAAAGCCC
351 ACATTGCTGG TTACATGTGT AAATCACTGC GTTATTGCTT TAGTCATTGT
401 CTCTATTTAG CAATGACAAG ACTGGAAGAA GTAAATAGAG AAGTGAACAT
451 GCATTCTTCA GTGCGGTATC TTGGCTATTT AGCCAGAATC AATTTATTGG
501 TTGCTATATG CTTAGGTCTA TACGTAAGAT GGGAAAAAAC AGCAAATTCC
551 TTAATTTTGG TAATTTTAT TCTTGGTCTT TTTGTCTTGG GAATCGCCAG
601 CATACCTCTAT TACTATTTT CAATGGAAGC AGCAAGTTTA AGTCTCTCCA
651 ATCTTTGGTT TGGATCTCTG CTTGGCCTCC TATGTTTTCT TGATAATTCA
701 TCCTTTAAAA ATGATGTAAA AGAAGAATCA ACCAAATATT TGCTTCTAAC
751 ATCCATAGTG TTAAGGATAT TGTGCTCTCT GGTGGAGAGA ATTTCTGGCT
801 ATGTCGCTCA TCGGCCCCACT TTACTAACCA CAGTTGAATT TCTGGAGCTT
851 GTTGGATTTC CCATTGCCAG CACAACATAG TTGGTGAGAG AGTCTCTGAG
901 TGTCATTTTG CTTGTGTAG CTCTGGCTAT GCTGATTATT GATCTGAGAA
951 TGAAATCTTT CTTAGCTATT CCAAACCTAG TTATTTTTCG AGTTTTGTGA
1001 TTTTTTTCCT CATTGGAAAC TCCCAAAAT CCGATTGCTT TTGCGTGTGT
1051 TTTTATTTGC CTGATAACTG ATCCTTTCCT TGACATTTAT TTTAGTGGAC
1101 TTTCAAGTAA TGAAAGATGG AAACCCCTTT TGTACCGTGG AAGAATTTGC
1151 AGAAGACTTT CAGTCGTTT TGCTGGAATG ATTGAGCTTA CATTTTTTAT
1201 TCTTTCCGCA TTCAAACCTA GAGACACTCA CCTCTGGTAT TTTGTAATAC
1251 CTGGCTTTTC CATTTTGGGA ATTTTCAGGA TGATTGTGCA TATTATTTT
1301 CTTTAACTC TTTGGGGATT CCATACCCAA TTAATGACTT GCCATAAAGT
1351 ATATTTTACT CACAGGACAG ATTACAATAG CCTTGATAGA ATCATGGCAT
1401 CCAAAAGGAT GCGCCATTTT TGCTTGATT CAGAGCAGTT GGTGTCTTT
1451 AGTCTTCTTG CAACAGCGAT TTTGGGAGCA GTTTCCTGGC AGCCAAACAA
1501 TGGAAATTTT TTGAGCATGT TCCTAATCGT TTTGCCATTG GAATCCATGG
1551 CTCATGGGCT CTTCCATGAA TTGGGTAAC TTTAGGAGG AACATCTGTT
1601 GGATATGCTA TTGTGATTCC CACCAACTTC TGCAGTCCTG ATGGTCAGCC
1651 AACACTGCTT CCCCAGAAC ATGTACAGGA GTTAAATTG AGGTCTACTG
1701 GCATGCTCAA TGCTATCCAA AGATTTTTT CATATCATAT GATTGAGACC
1751 TATGGATGTG ACTATCCAC AAGTGGACTG TCATTGATA CTCTGCATT
1801 CAACTAAAA GCTTTCCTCG AACTTCGGAC AGTGGATGGA CCCAGACATG
1851 ATACGTATAT TTTGTATTAC AGTGGGCACA CCCATGGTAC AGGAGAGTGG
1901 GCTCTAGCAG GTGGAGATAC ACTACGCCTT GACACACTTA TAGAATGGTG
1951 GAGAGAAAAG AATGGTTCCT TTTGTCCCG GCTTATTATC GTATTAGACA
2001 GCGAAAATTG AACCCCTTGG GTGAAAGAAG TGAGGAAAAT TAATGACCAG
2051 TATATTGCAG TGCAAGGAGC AGAGTTGATA AAAACAGTAG ATATTGAAAG
2101 AGCTGACCCG CCACAGCTAG GTGACTTTAC AAAAGACTGG GTAGAATATA
2151 ACTGCAACTC CTGTAATAAC ATCTGCTGGA CTGAAAAGGG ACGCACAGTG
2201 AAAGCAGTAT ATGGTGTGTC AAAACGGTGG AGTGACTACA CTCTGCATT
2251 GCCAACGGGA AGCGATGTGG CCAAGCACTG GATGTTACAC TTTCTCGTA
2301 TTACATATCC CTTAGTGCAT TTGGCAAATT GGTATGCGG TCTGAACCTT
2351 TTTTGGATCT GCAAACTTGG TTTTAGGTGC TTGAAAAGAT TAAAAATGAG
2401 TTGGTTTCTT CCTACTGTGC TGGACACAGG ACAAGGCTTC AAACCTGTCA
2451 AATCTTAATT TGGACCCCAA AGCGGGATAT TAATAAGCAC TCATACTACC
2501 AATTATCACT AACTTGCCAT TTTTGTATG CTGTATTTT ATTTGTGGAA
2551 AATACCTTGC TACTTCTGTA GCTGCTCTCA CTTTGTCTT TCTTAAGTAA
2601 TTATGGTATA TATAAGGCGT TGGGAAAAAA CATTTTATAA TGAAAGTATG
2651 TAGGGAGTCA AATGCTTACT GTAAATGCAT AAGAGACGTT AAAAAATAA
2701 CTGCACTTTC AGGAATGTTT GCTTATGGTC CTGATTAGAA AGAAACAGTT
```

```

2751 GTCTATGCTC TGCAATGGTC AATGATGAAT TACTAATGCC TTATTTTCTA
2801 GGCATATAAT AATAGTTTAG AGAATGTAGA CCAGATAAAT TTGTTTACTG
2851 TTTTAAGAAA ACTACCAGTT TACTTACAGA AGATTCTTTT TTCCAAACAG
2901 TAGGTTTCAT CCAAGACCAT TTGAAGAAGT GCAAACCTCT TCTCTTAGAA
2951 AAGAAAGAGG GCAGCCTAAA ATAAACGCAA AATTTGCTTA TACTCCATCA
3001 CATTGAGATG TCTTGGTTGT GACTTATTAC CAGTGTGGCA GAGAACCCAA
3051 GTTACATTTT AGATCAAAAT ATTCTTTATG TAGGTATTGT TAAAAGGCTA
3101 GAGCCTACAA GTTGCTCTTC CATGCGTTGG TCAGGGGGCC CTGAAAACAC
3151 TGGTAAATAT AAGAGTCTTT CTCAGGGTAA CTAAATGTTT TCTTAATGAA
3201 CAGTGTTTCC AGCTACAAAT TCTTCCAATA AATTGTCTTC CTTTTTGAAA
3251 AGTACTCTCA TAGAAGAAAT TTAGCAATTT CTCGTTGACT GACTCAGTCT
3301 ATTTTAAGTA TTCAGAAAAG ATTTTGATCC CCATTGAGTT AATGCTCTGC
3351 CTTGAAAATT ATTTTCTGA TCCTTGTTAG TGATAACATT TTTTCTCTAC
3401 TGAAGGTCAG AGGATAGGAA ACAAGTATTT CTCTTCTGGT ATACATGTAA
3451 TGTATCTCTG AAAAAAGTAT TCATATTGGC AATTTTAGTT AGGCATAATA
3501 TTGTGGTTGT AATTTTAAA ACTTAGTGTT TTGTCTGATT AAAGCAGGCA
3551 CTGATCAGGG TATCTCCTAA GAGGTAATTC ACTTCTTAT CTTTCCAAT
3601 AATTATTACA TTCTAAATTT TCATCTATGA GAAATAACAA ACAAGAAGGG
3651 AATAGAATTA AATTGGGGTA TAATCTAATC TTCATTGTTT AAATGGTTTG
3701 CCTTCTCACC ATTGAAGCCA TTTTCTTATA GCCTCAGAAA GAGGAAATAA
3751 TGCCCTCCACC ATTTTCTACC TGGTGACTTG AAAATTGAAC TTTTAAGTTA
3801 GGAAGAAAGT AGAGTCAGGG AACTTGATA CCACTATCTA TGCAGCATTG
3851 TTATAGTCTG ATTATTTCTG TGTGTTGAAT ATGATTTTCC TAATGCTCTA
3901 AATAAAATTT TGTAAAAAT CAAAAAATAA AAAAAAATAA CTTATCGATA
3951 CCGTCGACCT CGATGATGTC GAC

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 365 bp to 2455 bp; peptide length: 697
 Category: putative protein
 Classification: unset

```

1 MCKSLRYCFS HCLYLAMTRL EEVNRVNMH SSVRYLGYLA RINLLVAICL
51 GLYVRWEKTA NSLILVIFIL GLFVLGIA SI LYVYFSMEAA SLSLNLWFG
101 FLLGLLCFLD NSSPKNDVKE ESTKYLLLS IVLRILCSLV ERISGVVRHR
151 PTLLTTVEFL ELVGFAIAST TMLVEKSLSV ILLVVALAML IIDLRMKSFL
201 AIPNLVIFAV LFFSSLETPE KNPIAFACFF ICLITDPFLD IYFSGLSVTE
251 RWKPFLYRGR ICRRLSVVFA GMIELTFFIL SAFKLDRDTHL WYFVIPGFSI
301 FGIFRMICHI IFLLTLWGFH TKLNDCHKVY FTHRTDYNLSL DRIMASKGMR
351 HFCLISEQLV FFSLLATAIL GAVSWQPTNG IFLSMELIVL PLESMAHGLF
401 HELGNCLGGT SVGYAIVIPT NFCSPDGQPT LLPPEHVQEL NLRSTGMLNA
451 IQRFFAYHMI ETYGCYSTS GLSFDLHLSK LKAFLELRTV DGPRHDTYIL
501 YYSGHTHGTG EWALAGGDTL RDLTLIEWWR EKNGSFC SRL IIVLDSNST
551 PWVKEVRKIN DQYIAVQGA E LIKTVDIEEA DPPQLGDFTK DWVEYNCNSC
601 NNICWTEKGR TVKAVYGVS RWSDTLHLP TGSDVAKHWM LHFPRITYPL
651 VHLANWLCLG NLFWICKTCF RCLKRLKMSW FLPTVLDTGQ GFKLVKS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2c1, frame 2

PIR:A71148 hypothetical protein PH0395 - *Pyrococcus horikoshii*, N = 1,
 Score = 96, P = 0.12

>PIR:A71148 hypothetical protein PH0395 - *Pyrococcus horikoshii*
 Length = 288

HSPs:

Score = 96 (14.4 bits), Expect = 1.3e-01, P = 1.2e-01
 Identities = 59/234 (25%), Positives = 116/234 (49%)

Query: 77 IASILYFFSMEAASLSLSNLWFGFLL--GL--LCFLDNSSFKNDVKEESTKYLLLTSTIV 132
++ +LYY F+ A ++ L G+LL + L +L N + V+ + K + ++
Sbjct: 57 LSLVLYLFAFSALK-TIIFLALGYLLMNSIYELGYLMNDTISRVEGKVHKVRVKLTVF 115

Query: 133 LRILCSLVERISGYVRHRPTLLTTVEFLELVGFAIASTTMLVEKSLSVILLVVALAMLI 192
+L +L I YV ++ T+ FL+LVG ++ +L E +L ++ L+ L +
Sbjct: 116 DSSLIALSRAI--YV-----VIFTLVFLKLVLGLQYSTQVILAETVTLFLVFLLYDLTPKHV 168

Query: 193 DLRMKSFLAIPNLVIFAVLLFFSSLET-PKNPIAFACFFICLITDPFLDIYFSGLSVTER 251
M SF + + F +LL F T +N I + FI I F ++ + +
Sbjct: 169 RTVMLSF-PLKFMKAFVLLLPFIITGTLVENVITLS--FILPIAVRFSQAHYLTACKDN 225

Query: 252 WKPFYLRGRICRRLSVVFAGMIEL-TFFILSAFK-LRDTHLW-YFVIPGFSIFGIFRMIC 308
P ++ R+ R S+++ + L TF +L +F L +T L ++IP F++ + ++
Sbjct: 226 -PPRDFKRRV-ERFSMMYLQVTSLSFTVLVSFVYLGNTDLLRQYLIP-FAVNVVLILLS 282

Query: 309 HI 310
++
Sbjct: 283 YL 284

Pedant information for DKFZphfbr2_2c1, frame 2

Report for DKFZphfbr2_2c1.2

[LENGTH] 697
[MW] 79741.46
[pI] 8.41
[KW] TRANSMEMBRANE 11
[KW] LOW_COMPLEXITY 9.76 %

SEQ MCKSLRYCFSHCLYLAMTRLEEVNREVMHSSVRYLGYLARINLLVAICLGLYVRWEKTA
SEG
PRD cccceehhhhhhhhhhhhhhhhhhhhhccceehhhhhhhhhhhhhhhhhhhccccc
MEMMMMMMMMMMMMMMMMM.....

SEQ NSLILVIFILGLFVLGIASILYFFSMEAASLSLSNLWFGFLLGLLCLFLDNSSFKNDVKE
SEG ..xxxxxxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD cccceeeccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
MEM ...MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....

SEQ ESTKYLLLTSTIVLRILCSLVERISGYVRHRPTLLTTVEFLELVGFAIASTTMLVEKSLSV
SEGxxxxxxxxxxxxx.....xxxxxxxxxxxxx.....
PRD ccchhhhhhhhhhhhhhhhhhhhhccceccccchhhhhhhhhhhhhhhhhhhhhhhhh
MEMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....

SEQ ILLVVALAMLIIDLRMKSFLAIPNLVIFAVLLFFSSLET-PKNPIAFACFFICLITDPFLD
SEG xxxxxxxxxxxxxxxx.....
PRD hhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhcccccchhhhhhhhhccccc
MEM MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....

SEQ IYFSGLSVTERWKPFYLRGRICRRLSVVFAGMIELTFFILSAFKLRDTHLWYFVIPGFSI
SEG
PRD eeccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccceeecccc
MEMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....M

SEQ FGIFRMICHIIIFLLTLWGFHTKLNDCCHKVYFTHRTDYNLDRIMASKGMRHFCLISEQLV
SEG
PRD hhhhhhhhhhhhhhhhhhhhhcccccceeeccccccccchhhhhhhcccccchhhhhhh
MEM MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....MM

SEQ FFSLATAILGAVSWQPTNGIFLSMFLIVLPLESMAGLGFHELGNCLGGTSVGYAIVIPT
SEG
PRD hhhhhhhhhhhcccccchhhhhhhheehhhhhhhhhcccccceeecccc
MEM MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....

SEQ NFCSPDGPQTLLPPEHVQELNLRSTGMLNAIQRFAYHMIETYGCDYSTSGLSFDTLHSLK
SEG
PRD cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhcccccceccccchhhhh
MEM
SEQ LKAFLELRITVDGPRHDTYILYSGHTGTGEWALAGGDTLRDLTIEWWREKNGSFCSRL
SEG
PRD hhhhhhhhhcccccceeeccccccccceccccchhhhhhhhhhhcccccceee
MEM
SEQ IIVLDSNSTPWWKEVRKINDQYIAVQGAELIKTVDIEEADPPQLGDFTKDWVEYNCNSC

```

SEG .....
PRD eeeeeccccccchhhhhccceeeccccceeecccccccccccccccccccccccccc
MEM .....

SEQ NNICWTEKGRTVKAVYGVSKRWSYTLHLPTGSDVAKHWWMLHFPRITYPLVHLANWLCGL
SEG .....
PRD ceeeeccccceeeccccccccccccchhhhhcccccchhhhhhhhhcc
MEM .....

SEQ NLFWICKTCFRCLKRLKMSWFLPTVLDTGQGFKLVS
SEG .....
PRD eeeeehhhhhhhhhhhhccceeecccccccccc
MEM .....

```

(No Prosite data available for DKFZphfbr2_2c1.2)

(No Pfam data available for DKFZphfbr2_2c1.2)

DKFZphfbr2_2c17

group: signal transduction

DKFZphfbr2_2c17.3 encodes a novel 446 amino acid protein with similarity to yeast YMR131c and mammalian retinoblastoma-binding protein RbAp46

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to YMR131c and retinoblastoma-binding protein RbAp46

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2248 bp

Poly A stretch at pos. 2230, polyadenylation signal at pos. 2200

```
1 TGGGGAAGAT GGC GGCGCGCGC AAGGGTCGGC GTCGCACGTG TGA AACCGGG
51 GAACCCATGG AAGCCGAGTC CGGCGACACA AGTTCCGAGG GCCCGGCCCA
101 GGTCTACCTG CCCGGCCGGG GGCGCGCGCT ACGCGAAGGG GAGGAGCTGG
151 TCATGGACGA GGAGGCCCTAT GTGCTCTACC ACCGAGCGCA GACTGGCGCC
201 CCCTGTCTCA GCTTTGACAT AGTCCGGGAT CACCTGGGAG ACAACCGGAC
251 AGAGCTTCTT CTTACACTTT ACTTGTGTGC TGGGACCCAG GCTGAGAGCG
301 CCCAGAGCAA CAGACTGATG ATGCTTCGGA TGCACAATCT GCATGGGACA
351 AAGCCCCCAC CCTCAGAGGG CAGTGATGAA GAAGAAGAGG AGGAAGATGA
401 AGAGGATGAA GAAGAGCGGA AACCTCAGCT GGAGCTGGCC ATGGTGCCCC
451 ACTATGGTGG CATCAACCGA GTTCGGGTGT CATGGCTGGG TGAAGAGCCT
501 GTGGCTGGGG TGTGGTCAGA GAAGGGCCAG GTGGAGGTGT TTGCGCTGCG
551 GCGGCTTCTG CAGGTGGTGG AGGAGCCCCA GGCCCTGGCA GCCTTCCTCC
601 GGGATGAGCA GCGCCAAATG AAGCCCATCT TCTCCTTCGC TGGACACATG
651 GCGAGGGGCT TTGCCCTTGA CTGGTCCCCC CGGGTGACCG GTCGCTGCT
701 GACCGGTGAC TGTCAAAAGA ACATCCACCT CTGGACACCT ACGGACGGCG
751 GCTCCTGGCA CGTGGACCAG CGGCCATTCT TGGGCCACAC ACGCTCTGTG
801 GAGGACCTGC AGTGGTCACC GACTGAGAAC ACGGTGTTTG CCTCCTGCTC
851 AGCTGACGCC TCCATCCGCA TCTGGGACAT CCGGGCAGCC CCCAGCAAGG
901 CCTGCATGCT CACCACAGTC ACCGCCCATG ATGGGGACGT CAATGTCATC
951 AGCTGGAGCC GCCGGGAGCC CTTCTGTCTC AGTGGCGGGG ATGATGGGGC
1001 CCTCAAGATC TGGGACCTTC GGCAGTTCAA GTCTGGTTCC CCAGTGGCCA
1051 CCTTCAAGCA GCACGTGGCC CCCGTGACCT CCGTCGAGTG GCACCCCCAG
1101 GACAGCGGGG TCTTTGCAGC CTCGGGTGCA GACCACCAGA TCACACAGTG
1151 GGACCTGGCA GTGGAGCGGG ACCCTGAGGC GGGCGACGTG GAGGCCGACC
1201 CCGGACTGGC CGACCTCCCG CAGCAGCTGC TGTTCGTGCA CCAGGGCGAG
1251 ACCGAGCTGA AGGAGCTGCA CTGGCACCAG CAGTGCCAGG GGCTCCTGGT
1301 CAGCAGCGCG CTGTCAGGCT TCACCATCTT CCGCACCATC AGCGTCTGAG
1351 GCGTCCCACT GGCTCTGATC TTGCTTCCTG CTTGGAAACT GAAGTCGAAT
1401 TGGGCTCCCC TGGGAGGGGT TCATTGAGGT CTGTTGACTG AGACTGGCCG
1451 GCCTGTGGGC TGCCGTGATG GATTCTGTTT GACGTATTGT TCTCTAGAAG
1501 GCCTGGCTCT GATCCAGTGA CCCCTCTCAC CAAAGAACTC GGTTTAACCA
1551 GGGCTCTGTA AGACCACTCC CACCCAGAGA CTTGTGTGGC CTGGTGTGGC
1601 CTGTGTGTCG GATTCTTCC TGTCAGCTGT GACCCATTG ACCTGTGTCC
1651 CCAGAACCCA GTTTTTTGT TGTGTGTTG AGACGGAGTC TTGGTCTGTC
1701 GCCCAGGCTG GAGTGCACTA GCACGATCTT GGCTCACTGC AACCTCCGCC
1751 TCCTGGGTTA AAGTGATTCT CTCAGCTCAG TCTCCAGGT AGCTGGGATT
1801 ACAGGCATGT GCCACCACAC CCCGTTAATT TTTGTATTTT TAGTAGAGAC
1851 GGGGTTTTCAC CATGTTGGCC AGGCTGGTCT CAAATTCTTG ATCTCAAGTG
1901 ATCTGTCCCG CCCGGCCTCC CAGAGTGCTG GGTGGGATT ACAGGCGTGA
1951 GCCACCGCGT CCGGCTCAGG ACCCAGTTT GGCTGCTGGT TCCCAGCAGG
2001 GGAATCGGGG GATATACAGT GGCTGCACCA AATTGGAGGT GTGGGTTTCT
2051 CCAACACAAT TTGCTTCTGC CCGTTGTCTT CCTGCCAGCT GGGTTTGGCC
2101 AGGATTTCTC CGTGTGGGGG CTACATGCGA CCCTCTCCCC TCCTCCCTGA
2151 CTTTAGAGGC TGGTGCTGTG TCGGGAGGAA GGTGAGGCT CCTGAGCAGC
2201 AATAAAGGAC CAGGAAGAGG CCTGAGGTGG AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 9 bp to 1346 bp; peptide length: 446
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: WD_REPEATS (323-338)

```

1 MAARKGRRRT CETGEPMEAE SGTSSSEGA QVYLPGRGPP LREGEELVMD
51 EEAYVLYHRA QTGAPCLSF IVRDHLGDN TELPLTLYLC AGTQAESAQS
101 NRLMLRMHN LHGTPPPSE GSDEEEED EDEEEKRPQ LELAMVPHYG
151 GINRVRVSWL GEEPVGAVWS EKGQVEVFAL RRLQVVVEP QALAAFLRDE
201 QAQMKPIFSF AGHMGEGFAL DWSPRVTGRL LTGDCQKNIH LWTPTDGGSW
251 HVDQRPFVGH TRSVEDLQWS PTENTVFASC SADASIRIWD IRAAPSKACM
301 LTTVTADGDG VNVISWSRRE PFLLSGDDG ALKIWDLRQF KSGSPVATFK
351 QHVAPVTSVE WHPQDSGVFA ASGADHQITQ WDLAVERDPE AGDVEADPGL
401 ADLPQQLLFV HQGETELKEL HWHPQCPGLL VSTALSGFTI FRTISV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2c17, frame 3

TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence., N = 1, Score = 910, P = 2.7e-91

PIR:S53061 hypothetical protein YMR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 691, P = 4.3e-68

PIR:I49367 retinoblastoma-binding protein mRbAp46 - mouse, N = 1, Score = 338, P = 1.1e-30

PIR:I39181 retinoblastoma-binding protein RbAp46 - human, N = 1, Score = 338, P = 1.1e-30

>TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence.
 Length = 469

HSPs:

Score = 910 (136.5 bits), Expect = 2.7e-91, P = 2.7e-91
 Identities = 195/442 (44%), Positives = 259/442 (58%)

```

Query: 18 EAESGDTSSSEGAQVYLPGRGPPLREGEELVMD E EAYVLYHRAQTGAPCLSF DIVRDHLG 77
      EA S + S P + V + PG L + GEEL D AY H G PCLSF DI + D LG
Sbjct: 18 EASSSEIPSI-PTRVWQPGVDT-LEDGEELQCDPSAYNSLHG FHVGVWPCLSFDILGDKLG 75

Query: 78 DNRTELPLTLYLCAGTQAESAQSNRLMLRMHNLHGTPK---PPSEGSDEEEED- 133
      NRTE P TLY+ AGTQAE A N + + ++ N+ G + P + G+ E+E+E+DE+D
Sbjct: 76 LNRTEFPHTLYMVAGTQAEKAAHNSIGLFKITNVSGKRRDVPKTFNGEDEDEDED 135

Query: 134 -----EEERKPQLELAMVPHYGGINRVRVSWLGEEPVAGVWSEKQVEVFALRRLQ 185
      E + P.+++ V H+G +NR+R + W++ G V+V+ + L
Sbjct: 136 DSDDDGDGEASKTPNIQVRRVAHGCVCNIRAMPQNSH-ICVSWADSGHVQVWDMSSHLN 194

Query: 186 VVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSPRVTGRLLTGDCQKNIHLWTPT 245
      + E + P+ +F+GH EG+A+DWSP GRLL+GDC+ IHLW P
Sbjct: 195 ALAESETEGKDGTSFVLNQAPLVNFSGHKDEGYAIDWSPATAGRLLSGDCCKSMIHLWEPA 254

Query: 246 DGGSWHVDQRPFVGHTRSVEDLQWSPPTENTVFASCSADASIRIWDIRAAPSKACMLTTVT 305
      G SW VD PF GHT SVEDLQWSP E VFASCS D S+ +WDIR S A +
Sbjct: 255 SG-SWAVDPIPFAGHTASVEDLQWSPAENVFASCSVDGSAVAVDIRLGKSPAL---SFK 310

Query: 306 AHDGDVNVISWSRREPFL-SSGDDGALKIWDLRQFKSGSPV-ATFKQHVAPVTSVEVHP 363
      AH+ DVNVISW+R +L SG DDG I DLR K G V A F+ H P+TS+EW
Sbjct: 311 AHNADVNVISWNRLASCLASGDDGTFSIRDLRLIKGGDAVVAHFEYHKHPITSIEWSA 370

```

Query: 364 QDSGVFAASGADHQITQWDLAVERDPE-----AGDVEADPGLADLPQQLLFVHQGETEL 417
 ++ A + D+Q+T WDL++E+D E A E DLP QLLFVHQG+ +L
 Sbjct: 371 HEASTLAVTSGDNQLTIWDLSEKDEEEAEFNAQTKELVNTPDLPQQLLFVHQGQKDL 430

Query: 418 KELHWHWPQCPGLLVSTALSGFTIFRTISV 446
 KELHWH Q PG+++STA GF I ++
 Sbjct: 431 KELHWHNQIPGMIISTAGDGFNIILPYNI 459

Pedant information for DKF2phfbr2_2c17, frame 3

Report for DKF2phfbr2_2c17.3

[LENGTH] 446
 [MW] 49447.38
 [pI] 4.82
 [HOMOL] TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein";
 Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence. 1e-90
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YMR131c] 4e-65
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YPR178w] 1e-11
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-11
 [FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 4e-09
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 4e-09
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
 YDL145c] 5e-09
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c]
 5e-09
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c
 TAF90 - TFIID subunit] 6e-09
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YMR116c] 5e-08
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 5e-08
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YLR429w] 3e-07
 [FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR142c] 3e-06
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c]
 3e-06
 [FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YDR142c] 3e-06
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 4e-06
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YER107c] 4e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YKL021c] 4e-06
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YER107c] 4e-06
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 2e-05
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c]
 2e-05
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YIL046w]
 2e-05
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YIL046w] 2e-05
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YLL011w] 3e-05
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YOR212w] 5e-05
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YOR212w] 5e-05
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YOR212w] 5e-05
 [BLOCKS] BL00678
 [SCOP] d2trcb_2.51.3.1.1 Transducin (heterotrimeric G protein), gamma 5e-29
 [PIRKW] plasma 6e-07
 [PIRKW] duplication 4e-12
 [PIRKW] hormone 6e-07
 [PIRKW] transmembrane protein 1e-07
 [PIRKW] stomach 6e-07
 [PIRKW] actin binding 1e-07
 [PIRKW] leucine zipper 1e-07
 [PIRKW] signal transduction 2e-06
 [PIRKW] heterotrimer 2e-06
 [PIRKW] peripheral membrane protein 6e-07
 [PIRKW] GTP binding 2e-06
 [SUPFAM] WD repeat homology 1e-63
 [SUPFAM] yeast coatomer complex alpha chain 1e-07
 [SUPFAM] GTP-binding regulatory protein beta chain 4e-07
 [SUPFAM] PRL1 protein 8e-09

[SUPFAM] MS11 protein 4e-12
 [SUPFAM] coatomer complex beta' chain 1e-09
 [PROSITE] WD_REPEATS 1
 [PFAM] WD domain, G-beta repeats
 [KW] All_Beta
 [KW] 3D
 [KW] LOW_COMPLEXITY 3.14 %

SEQ MAARKGRRRTCTETGEPMEAESGDTSSSEGPAQVYLPGRGPPLREGEELVMDEEAYVLYHRA
 SEG
 lgotB
 SEQ QTGAPCLSFDIVRDHLGDNRTPLTLTYLCAGTQAESAQSNRLMLRMHNLHGTKPPSE
 SEG
 lgotB
 SEQ GSDEEEEEDEEDEERKPQLELAMVPHYGGINRVVSWLGEVPVAGVWSEKQVEVFAL
 SEG ..xxxxxxxxxxxxxxxx.....
 lgotB
 SEQ RRLQVVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSPRVTGRLLTGDCQKNIH
 SEG
 lgotBEECCCCCEEEEEETTT-TCEEEEETTTEEE
 SEQ LWTPTDGGSWHVDQRPVGHTRSVEDLQWSPTENTVFASCSADASIRIWDIRAAFSKACM
 SEG
 lgotB EEETTTT---CEEEEECCCCCEEEEEETTCE-EEEEETTTEEEEEETTT--TEEEE
 SEQ LTTVTAHDGVDNVISWSRREPFLSGGDDGALKIWDLRQFKSGSPVATFKQHVAPVTSVE
 SEG
 lgotB EECBTTBTCEEEEEETTTTTEEEEEETTTEEEEE.....
 SEQ WHPDQSGVFAASGADHQITQWDLAVERDPEAGDVEADPGLADLPQQLLFVHQGETELKEL
 SEG
 lgotB
 SEQ HWHPQCPGLLVSTALSGFTIFRTISV
 SEG
 lgotB

Prosites for DKFZphfbr2_2c17.3

PS00678 323->338 WD_REPEATS PDOC00574

Pfam for DKFZphfbr2_2c17.3

HMM_NAME WD domain, G-beta repeats
 HMM *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
 ++GH+ V ++ +SP + +++S S D ++R+WD
 Query 257 FVGHTRSVEDLQWSPTENTVFASCSADASIRIWD 290
 24.88 304 336 1 34 dkfzphfbr2_2c17.3 similarity to YMR131c and retinoblastoma-
 binding protein RbAp46
 Alignment to HMM consensus:
 Query *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
 + H+++V+ +++S + ++SG++DG +++WD
 dkfzphfbr2 304 VTAHDGVDNVISWSRREPFLSGGDDGALKIWD 336

DKFZphfbr2_2c18

group: brain associated

DKFZphfbr2_2c18 encodes a novel 302 amino acid protein with weak similarity to cyclin-dependent kinase p130-PITSLRE.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to cyclin-dependent kinase p130-PITSLRE

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2835 bp

Poly A stretch at pos. 2817, polyadenylation signal at pos. 2796

```
1 TGGGGCGGAC GCGGAGGGAG TCCAGAGCCT TGAGCCCGGT GCTCCTCCCT
51 CGCGCAGCGG TGGCTCTGCG GCCGCTGGAG TAAACACTGC CTTTGTTCCC
101 TAGCGCCTCG TCTTTCGTCTG CCCCCTGCCG TCACGCCGCC GGGCTCTGGC
151 CGGCCCCGCC TCGGTCCCTTG AACCCCATTT CGGCTCGTGC CGTGCGGATG
201 CAGCTGCCGG GCCTGGGTIT GGGCATTGAG CGGGAGGAGG AGGAGGAGCG
251 GCGGCGCCTG GCGGCGCATGC GATGGGGAAC TGCTGCTGGA CGCAGTGCTT
301 CGGACTGCTT CGCAAGGAAG CGGGGCGGCT GCAGCGAGTA GCGGCGGCGC
351 GAGGATCCAA GTATTTTAGA ACATGCTCAA GAGGTGAGCA CTTGACAATA
401 GAGTTTGAAG ATCTAGTAGA AAGTGATGAA GGGGAGAGCC CAGGAAGCAG
451 TCATAGGCCT CTTACTGAGG AAGAAATTGT TGACCTAAGA GAAAGGCATT
501 ATGATTCAT TGCCGAAAAA CAAAAAGATC TTGATGAGAA AATTCAAAAA
551 GAGTTAGCCT TACAAGAAGA GAAGTTAAGA CTAGAAGAAG AAGCTTTATA
601 CGCTGCACAG CGTGAAGCAG CCAGGGCAGC AAAGCAGCGA AAGCTCTTGG
651 AGCAAGAAAG GCAGAGAATT GTGCAGCAAT ATCATCCTTC CAACAATGGA
701 GAATATCAAA GTTCAGGACC AGAAGATGAC TTCGAATCTT GTTTGAGAAA
751 TATGAAGTCA CAGTATGAAG TTTTTCGAAG TAGTAGACTC TCATCAGATG
801 CTACAGTTT GACACCAAAT ACAGAAAGCA GTTGTGATTT AATGACCAAA
851 ACTAAATCAA CTAGTGGAAA TGACGACAGC ACATCCTTAG ATCTAGAGTG
901 GGAAGATGAA GAAGGAATGA ATAGAATGCT TCCAATGAGA GAACGTTCCA
951 AAACAGAGGA AGACATTCTA CGGGCAGCAC TTAAGTATAG CAACAAGAAG
1001 ACTGGAAGTA ATCTACATC AGCCTCTGAT GATTCCAATG GGCTGGAGTG
1051 GGAATATGAT TTTGTTAGTG CCGAAATGGA TGATAATGGA AATTCGAGT
1101 ATCTTGGATT TGTAATCCT GTATTAGAAC TGTCTGATTC TGGCATAAGG
1151 CATTCTGACA CAGATCAACA GACTCGATAG GGTAAATTTG TGTGACCTTG
1201 TTTATCAGTT ATGACCAAAT GTTAAAAACC AACTAGAATG TATAAGTGAT
1251 TGTGCTTAGC CTTTTGTAA GGGAGATGTG TAAGAAACCA TGCTGTAAT
1301 GCTTATTTTA TTACAAAAGG GTAGGGATGA TAGGATCTGA ATTGATACAG
1351 TTTCAAGTGC AATTTCATCA TCTGCCTTCT GCTTTTCAAG ACCAATTTAA
1401 TGGTCTGTGC ATGTTACTGA TTAAATTTAC TTTGCTTGT CTTTATAGCA
1451 TTTCTGTTTA CTATGGTAGA TTTCCACTTT CAATTTTAA AATTAATTTT
1501 ATTTTGAATG ATTTATGAAG CCTATTTTCA TGTCTAACTA TGAATAATTT
1551 AAGACTTTTT TGTTAATTCT CAGCCGATGT GAAGGAAGCA TGAGGAGGGA
1601 TCGTCAGACT CAGATTTAGA ATAGTGTTCC CGTTTCCAGC ATTATTTAT
1651 TCTATGACTT CTTTGGATTT TATTATCTAA TAGTAAGTAC AGTTGATGTG
1701 GGTAGATGAC TCTAAGAAAT GCTGAAGTAT CGGCATTACA TGTGTTTAT
1751 TACATGTCTT AGTTTGATAA TGTGATTCA ATCTGAACAA AAGATAATAT
1801 AAAAATAACC CTTCAGAGTT TGGACATTTC AAGTTGGTAA TAATAAAAAA
1851 TAATATTTAA GAAGATATAT ATATATATAT ATTTAGTTT TTCCACTTCA
1901 TTTTACATGC CACTATATTG ACTTTAATTG ATATACAGTA TTAAGTTTTT
1951 AGGTGCCATT ATTTTAAAAA AATTCTATAT TTCCAATGAA CGATGTTAGA
2001 TTTTACACAG AACATATTCT CTGCATGATT TCAGAAAAGA AAATCTAAAA
2051 AGGTAATACG GGTATTTCAA ATAAAAATCCT TTCTGGTATG AAAGGCTCCA
2101 TTGATTTTAT TAAGCCTTCC TTTACCTTGT AGTACAAGGT GCTTTAATGG
2151 GATAGAATTA AGCATATCAA TATCTATAAC TGCAATTTGT GCTAGACAAT
2201 TACTGTCTCT TTCTCTAAAA TGTATATGTC AATTTACAAG GCCAGGGATA
2251 GAAAAACATC CATAATTGCT TTCCTTGATT TTGCTGAGGA TTTGGTATGA
2301 TTTTAGTAGA CAACTGTTT TTTGGTTTTT CCTTAATGTT TTTAATTTT
2351 TTTCTCTCTG CAACAATGAC GGTGCATGTT CTTATAAATA TAGGAAGGTC
2401 CAGATATAAA TAGTAACCTA AAGTCTTTCG TGTGCTTAAA AAAAAAATC
2451 ATGTGGCTCT TTCAATATTT GAACTGCTAA GCAATGACAT CTGTAGTTTT
2501 ATCTCCTTTT TTATGTCATA GAAATTAATA TGATACTTTA AATATGTAAG
2551 TATAATACAT TGGTAATGCT ATTATTTATA TCTGTCTTAA CATAATTTAA
2601 GTTGTAGCTG TGTCTTGGAA ATATTTTAA GGTAACTAT ATTCACATTG
2651 CCTGTGTAA TGCTTTTAA GGTGTGTATA CATCAGATGT ATATTTTGG
```

2701 TTTGGCATAA GCTACGATTG TAATTTTCT TGGCTTTTG TTCATAAAGA
 2751 ATTTTGTGAA GGAATGGTAA CAAATGGTAA TTTACAAATG GTTGTGAATA
 2801 AACACATTTT TACACTTAAA AAAAAAAAAA AAAAA

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 272 bp to 1177 bp; peptide length: 302
 Category: similarity to known protein

1 MGNCWTQCF GLLRKEAGRL QRVGGGGGSK YFRTCSRGEH LTIEFENLVE
 51 SDEGESPGSS HRPLTEEEIV DLRERHYDSI AEKQKDLDEK IQKELALQEE
 101 KLRLEEEALY AAQREAAARA KQRKLEQER QRIVQQYHPS NNGEYQSSGP
 151 EDDFESCLRNMKSQYEVFRS SRLSSDATVL TPNTSSCDL MTKTKSTSGN
 201 DDSTSLDLEW EDEEGMNRML PMRERSKTEE DILRAALKYS NKKTGSNPTS
 251 ASDDSNGLEW ENDFVSAEMD DNGNSEYSGF VNPVLELSDS GIRHSDTDQQ
 301 TR

BLASTP hits

Entry A55817 from database PIR:
 cyclin-dependent kinase p130-PITSLRE - mouse
 Length = 783
 Score = 123 (43.3 bits), Expect = 0.00013, P = 0.00013
 Identities = 53/197 (26%), Positives = 96/197 (48%)

Alert BLASTP hits for DKFZphfbr2_2c18, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2c18, frame 2

Report for DKFZphfbr2_2c18.2

[LENGTH] 302
 [MW] 34281.39
 [pI] 4.73
 [PROSITE] MYRISTYL 5
 [PROSITE] CK2_PHOSPHO_SITE 12
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 3
 [KW] All_Alpha
 [KW] LOW_COMPLEXITY 13.58 %
 [KW] COILED_COIL 13.58 %

SEQ MGNCWTQCFGLLRKEAGRLQRVGGGGGSKYFRTCSRGEHLTIEFENLVESDEGESPGSS
 SEGxxxxx.....
 PRD cccccccchhhhhhhhhheeeccccccccceccccccchhhhhhhcccccccccc
 COILS
 SEQ HRPLTEEEIVDLRERHYDSIAEKQKDLDEKIQKELALQEEKLRLEEEALYAAQREAAARA
 SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
 PRD ccchhhhhhhhhccchhh
 COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
 SEQ KQRKLEQERQRIVQQYHPSNNGEYQSSGPEDDFESCLRNMKSQYEVFRSSRLSSDATVL
 SEG xxxxxxxx.....
 PRD hhhhhhhhhhhhhhhccccccccccccccccchhhhhhhhhheeeccccccceee
 COILS CCCCCCCC.....

```

SEQ      TPNTSSCDLMTKTKSTSGNDDSTSLDLEWEDEEGMNRMLPMRERSKTEEDILRAALKYS
SEG      .....
PRD      cccccccccccccccccccccccccchhhhhhhccccccccchhhhhhhcchhhhhhhhhhhc
COILS    .....

SEQ      NKKTGSNPTSASDDSNLEWENDFVSAEMDDNGNSEYSGFVNPVLELSDSGIRHSDDTDQQ
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....

SEQ      TR
SEG      ..
PRD      CC
COILS    ..

```

Prosites for DKFZphfbr2_2c18.2

PS00005	60->63	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00006	36->40	CK2_PHOSPHO_SITE	PDOC00006
PS00006	65->69	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	198->202	CK2_PHOSPHO_SITE	PDOC00006
PS00006	204->208	CK2_PHOSPHO_SITE	PDOC00006
PS00006	226->230	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	245->251	MYRISTYL	PDOC00008
PS00008	291->297	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2c18.2)

DKFZphfbr2_2d15

group: differentiation/development

DKFZphfbr2_2d15 encodes a novel 438 amino acid protein similarity to Mus musculus testis-specific Y-encoded-like protein (Tspyl1).

The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. The novel protein is a new member of the TSPY-SET-NAP1L1 family, which represents proteins closely related to TSPY. Therefore, the new protein seems to be involved in early spermatogenesis.

The new protein can find application in modulating early spermatogenesis.

strong similarity to testis-specific Y-encoded-like protein

complete cDNA, complete cds, EST hits
localisation: primer B does not match perfect

Sequenced by Qiagen

Locus: /map="729.2 cR from top of Chr6 linkage group"

Insert length: 3229 bp

Poly A stretch at pos. 3206, polyadenylation signal at pos. 3184

```
1 GGAGACTGTA GGGTGGGCGG TGCAGCGGC GGTAGCTCC CAGTTCGGCC
51 TCTGAGGAAA ACGGGCGTTC GCCTGCGGTT GGTCGCACTG TTAGCAACAT
101 GAGCGGCTCG GATGGGTCA AGAGGACCAC TCCCTCCAA ACCCACAGCA
151 TCATTATTTG TGACCAAGTC CCGAGCGACC AGGACGCACA CCAGTACCTG
201 AGGCTCCGCG ACCAAAGCGA GGCGACACAG GTGATGGCGG AGCCGGGTGA
251 GGGAGGCTCG GAGACCGTCG CGCTCCCGCC TTCACCGCCT TCAGAGGAGG
301 GGGCGGTACC CCAGGATCCC GCGGCGGTG GCGGTACTCC CCAGATCCGA
351 TGTGTTGGGG GTCGCGGTCA TGTGGCGATC AAAGCCGGGC AGGAAGAGGG
401 CCAGCCTCCC GCCGAAGGCC TGGCAGCCGC TTCTGTGGTG ATGGCAGCCG
451 ACCGCAGCCT GAAAAAGGGC GTTCAGGGTG GAGAGAAGGC CCTAGAAATC
501 TGTGGCGCCC AGAGATCCGC GTCTGAGCTG ACGGCGGGGG CGGAGGCTGA
551 GGGCGAGGAG GTGAAGACAG GAAAGTGC GC CACCTCTCA GCAGCCGTGG
601 CTGAGAGGGA GAGCGCTGAG GTGGTGGTGA AGGAAGGCCT GGCAGAGAAG
651 GAGGTAATGG AGGAGCAGAT GGAGGTAGAG GAGCAGCCGC CAGAAGGTGA
701 AGAAATAGAA GTGGCGGAGG AGGATAGATT GGAGGAGGAG GCGAGGGAGG
751 AAGAAGGGCC CTGGCCTTTG CATGAGGCTC TCCGCATGGA CCCTCTGGAG
801 GCCATCCAGC TGGAACTGGA CACTGTGAAT GCTCAGGCCG ACAGGGCCTT
851 CCAACAGCTG GAGCACAAGT TTGGGCGGAT GCGTCGACAC TACCTGGAGC
901 GGAGGAACCTA CATATTTCAG AATATCCCGG GCTTCTGGAT GACTGCTTTT
951 CGAAACACCC CCCAGTTGTC CGCCATGATT AGGGGCCAAG ATGCAGAGAT
1001 GTTAAGGTAC ATAACCAATT TAGAGGTGAA GGAACCTAGA CACCTAGAA
1051 CCGGTTGCAA GTTCAAGTTC TTCTTTAGAA GAAACCCCTA CTTCAGAAAC
1101 AAGCTGATTG TCAAGGAATA TGAGGTAAGA TCCTCCGGCC GAGTGGTGTC
1151 TCTTCTTACT CCAATTATAT GGCGCAGGGG GCATGAACCC CAGTCCCTCA
1201 TTCGAGAGAA CCAAGACCTC ATCTGCAGCT TCTTCACTTG GTTTTCAGAC
1251 CACAGCCTTC CAGAGTCCGA CAAAATTGCT GAGATTATTA AAGAGGATCT
1301 GTGGCCAAAT CCACTGCAAT ACTACCTGTT GCGTGAAGGA GTCCGTAGAG
1351 CCCGACGTCG CCCGCTAAGG GAGCCTGTAG AGATCCCCAG GCCCTTTGGG
1401 TTCCAGCTCG GTTAACATTT GCCCTTGGGA ATACTCTGCG ACAAGGTCTC
1451 CTACACACCTT CTGCTGGACC TGTGCTTGGG CATCAGCAAT GAGTATGCCT
1501 TCTATTGTGC TTTGTTTTTG CTGACTTTTC TGCACCCTGT TTCCTTTGGA
1551 TATTCACTTC TCTCAACCTC AAGATTGAGA CGGTGGTGGG TATGCTTCTC
1601 CACTTCCATA TGACCTTCAT GCTGTCTGGA AATATCACAT GCTACGAGGT
1651 CATCTTCAC ACTACTTGTA AGCCAAGCAA ATGATACTGT AGATTGTACT
1701 GCCTTTATCT GCACTGCTTG GACCCTGTTT ATTTCCAGGG CCTCTGAAC
1751 GGTGCTGTCT ACTTGGATTT CTAGCTTTGG GAGCCTGTTC CACCTACTCA
1801 GCTCTGCATT GAGCAGTATG GGCACATGCC CTGTGGACAG TTAAGGACG
1851 TTAATGAACT CAGAGGAGAA AAGCAGTGAG CCACTTGTTT TGTGTGATTT
1901 ATGGTACTTC ATTGCTCTTC CTTACCTCTC AGTCACTTTC TATTGCTACC
1951 TGCCCTACAT TGGCTCTGCG CAAGGTCCCT CTCTCTCCCT GTTTTCTCTT
2001 TTTTTTTTTT TTTTTTTTTT TTTTGAGACG GAGGACGGAG TCTTGCTCTG
2051 TCGCCAGGTG TGGAGTGCAG TGGCGCGATC TCGGCTCACT GCAACCTCCA
2101 CCTCCCGGGT TCAAGCGATT CTCTGCTCCT AGCCTCCCGA GTAGCTGGGA
2151 CTACAGGCGC GCGCCGCCAC GCCCGGCTAA TTTTATATT TTTAGTAGAG
2201 ACGGGGTTTC ACCATGCTGG CCAGGCTGGT CTCGAACCCC GACCTCGTGA
2251 TCCGCCCTCC TTAGCCTCCC AATCCTCTCT TAAAAAAGTG ATAGCTCAGA
2301 AATATTTGTA AAAGCAAGGT TTTTATTTCA TTTTGGCTCT GTCATTTTCA
2351 GAGGCAAGA AGTTGGCCTG TAAAAAGAG TGCTAGAGCT CTTACGCCCC
2401 TCCCTTCTT CCCAACTTCC TACTTCTTAG CCCTTTTATC AACTCCTAGA
2451 ATAGTTAAAG AGAGACACAT CTAGATGGGA TGAAAGGTGC CCTAAGCAGG
```



```

2501 AGAACTGAA CAAAAGGCTA GAGGCATGGG CCAGGTAAAA ATTGGGCCTA
2551 GAGTGAAGAC TGTGCTGCCG TTAAGAGCTT TCGAGGAAGG AGTACTTACT
2601 CCCCAATGAT GATGAATGGA GAAATACTTT TCAGGGAGAA TTGAAGGGGT
2651 TAAAGTGTTA AATATGTTGC CTAGACAAGG GTTCTTTAAA GAAAGACAGC
2701 GCAACTTTGA ATGCTTCTT ACTTGTTTGG TGACCTAATT TATGTGGAAG
2751 ATTGTTATTT CATTAGGATT TAGTAAAATT TTTTCTCTG ATTCTAAACT
2801 TATTGTGAAA ATTGAGCTGT ACAGATATTC TTTTGATTTC AATTGGGAAC
2851 ATTTGGAAGA ACAACAGTCT TACTTGCCCTG TACAATATAG AGACATATGA
2901 ATAGTCATAA CAGTTTTCAG CTGTGCTCTG TTTCTGTAA ACTATATTCC
2951 TAGAAACATA GTTTGAACAA CTGGTCTTT GTTAGGCTTG TCAAATTGCC
3001 TTCATGGAAA AATAATCTAC AAAAGTATGG TTTAATTGAT TGTCTTACAT
3051 GATAATTTTC CCTGGCAACA ACTTAGTAAG TGATATATCT TTTTCTCTAA
3101 ATTGCTTAAA TACTGTGAAA TTGCTCTGAC AAATTGGAAG TGTACCATTG
3151 GCATATTTGT CTTCTTTTT ATGCATGATG GTAAAATAAA AGCATGTTGT
3201 TCTGCTAAGA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry AF042181 from database EMBLNEW:
Homo sapiens testis-specific Y-encoded-like protein (TSPYL) mRNA,
partial cds.
Score = 3411, P = 6.9e-148, identities = 685/687

Entry HS938343 from database EMBL:
human STS WI-11947.
Score = 1195, P = 2.1e-46, identities = 273/299

Medline entries

98399864:
Murine and human TSPYL genes: novel members of the TSPY-SET-NAP1L1 family

Peptide information for frame 3

ORF from 99 bp to 1412 bp; peptide length: 438
Category: strong similarity to known protein
Classification: Differentiation/Development

```

1 MSGLDGKVRT TPLQTHSIII SDQVPSDQDA HQYLRLRDQS EATQVMAEPG
51 EGGSETVALP PSPPSEEGGV PQDPAGRGGT PQIRVVGGRG HVAIKAGQEE
101 QPPAEGALAA ASVVMAADRS LKKGVGQGEK ALEICGAQRS ASELTAGAEA
151 EAEEVKTKGC ATVSAAVAER ESAEVVVKEG LAEKEVMEEQ MEVEEQPPEG
201 EEIEVAEEDR LEEEAAREEG PWPLHEALRM DPLEAIQLEL DTVNAQADRA
251 FQQLHKKFGR MRRHYLERRN YIIQNI PGFW MTAFRNHPQL SAMIRGQDAE
301 MLRYITNLEV KELRHPRTGC KEKFFFRNRP YFRNKLIVKE YEVRSRGRVV
351 SLSTPIIWRG GHEPQSFIIR NQDLICSFRT WFSDHSLPES DKIAEIIKED
401 LWPNPQLQYL LREGVRRARR RPLREPVEIP RPFQFQSG

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d15, frame 3

TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds., N = 1, Score = 1202, P = 3.1e-122

TREMBL:AB018264_1 gene: "KIAA0721"; product: "KIAA0721 protein"; Homo sapiens mRNA for KIAA0721 protein, partial cds., N = 1, Score = 798, P = 2e-79

TREMBL:AB015345_1 gene: "HRIHFB2216"; Homo sapiens HRIHFB2216 mRNA, partial cds., N = 1, Score = 570, P = 2.9e-55

>TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds.
Length = 379

HSPs:

Score = 1202 (180.3 bits), Expect = 3.1e-122, P = 3.1e-122
Identities = 258/377 (68%), Positives = 283/377 (75%)

```

Query:   62 SPPSEEGVVPQDPAGR-----GGTPQIRVVGGRGHVAIKAGQEE--GQP-P--AEGLA 110
          SP +EG   D   G           GTP R + G           G+   G P P   EGL
Sbjct:   3 SPERDEGTPVPDSRGHCADTVSGTPDRRPLLGEKAVTGEGRAGIVGSPAPRDVEGLVP 62

Query:  111 ASVVMADRSLLK-KGVQGGKEKALEICGAQRSASELTAGAEAEAEVKTGKCATVSAVAE 169
          V AA +   V+G   A+ +   ++ T GAE++A +VKT +   TV+AA
Sbjct:  63 QIRVAAARQGESPSPSVRGPAAAVFVTPKYVEKAQETRGAESQARDVKT-EPGTVA--- 119

Query:  170 RESAEVVVKEGLAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGWPWPLHEALR 229
          E +EV           EE MEVE Q P GEE+E+ E   EA EE GPW L   LR
Sbjct:  120 -EKSEVATPGS-----EEVMEVE-QKPAGEEMEMLEASGGVREAPPEAGPWHLGIDLR 170

Query:  230 MDPLEAIIQLELDTVNAQADRAFFQLEHKFGRMRHYLERNNYIIQNIPIGFWMATAFRNHPQ 289
          +PLEAIIQLELDTVNAQADRAFFQ LE KFGMRMRHYLERNNYIIQNIPIGFWMATAFRNHPQ
Sbjct:  171 RNPLEAIIQLELDTVNAQADRAFFQLEHKFGRMRHYLERNNYIIQNIPIGFWMATAFRNHPQ 230

Query:  290 LSAMIRGQDAEMLRYITNLEVKELRHPRTGCKFKFFFRNNPYFRNKLIIVKEYEVRSSGRV 349
          LSAMIRG+DAEMLRY+T+LEVKELRHP+TGCKFKFFFRNNPYFRNKLIIVKEYEVRSSGRV
Sbjct:  231 LSAMIRGRDAEMLRYVTSLEVKELRHPKTGCKFKFFFRNNPYFRNKLIIVKEYEVRSSGRV 290

Query:  350 VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSDSLPSDKIAEIIKEDLWPNPLQYY 409
          VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSDSLPSDKIAEIIKEDLWPNPLQYY
Sbjct:  291 VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSDSLPSDKIAEIIKEDLWPNPLQYY 350

Query:  410 LLREGVRRARRRPLREPVEIPRPFQSG 438
          L REG+RR RRRP+REPVEIPRPFQSG
Sbjct:  351 LCREGIRRRRRPIREPVEIPRPFQSG 379

```

Pedant information for DKFZphfbr2_2d15, frame 3

Report for DKFZphfbr2_2d15.3

```

[LENGTH]      438
[MW]           49307.65
[PI]           5.36
[HOMOL]        TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like
protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds. 1e-
107
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR048c]
1e-07
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKR048c] 1e-07
[BLOCKS]       BL00376F
[PIRKW]        nucleus 6e-39
[PIRKW]        DNA binding 3e-06
[PIRKW]        phosphoprotein 6e-39
[PIRKW]        alternative splicing 6e-39
[KW]           Alpha Beta
[KW]           LOW_COMPLEXITY 22.83 %

```

```

SEQ  MSGLDGVKRTTLPQTSHIIISDQVPSDQDAHQYLRLRDQSEATQVMAEPGEGGSETVALP
SEG  .....X
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  PSPPSEEGVVPQDPAGRGTPQIRVVGGRGHVAIKAGQEEGQPPAEGLAAASVVMADRS
SEG  xxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  LKKGVGQGGKEKALEICGAQRSASELTAGAEAEAEVKTGKCATVSAVAERESAEEVVVKEG
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  LAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGWPWPLHEALRMDPLEAIIQLEL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  DTVNAQADRAFFQLEHKFGRMRHYLERNNYIIQNIPIGFWMATAFRNHPQLSAMIRGQDAE
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ  MLRYITNLEVKELRHPRTGCKFKFFFRNNPYFRNKLIIVKEYEVRSSGRVSLSTPIIWR

```

```
SEG .....
PRD hhhhhhhhhhhhhccccceeeeeccccccchhhhhccccccccccccceeeccc

SEQ GHEPQSFIRRNQDLICSFFTWFSDHSLPESDKIAEIIKEDLWPNPLQYYLLREGVRRARR
SEG .....xxxxxxxxxxx
PRD cccccchhhhhccccceeeeeccccccchhhhhhhhhccccceeeccccchhhh

SEQ RPLREPVEIPRPFQSG
SEG xxxxxxxx.....
PRD hcccccccccccccccc
```

(No Prosite data available for DKF2phfbr2_2d15.3)

(No Pfam data available for DKF2phfbr2_2d15.3)

DKFZphfbr2_2d17

group: transmembrane proteins

DKFZphfbr2_2d17 encodes a novel 292 amino acid protein with similarity to a C.elegans hypothetical protein.

One transmembrane region is predicted for the protein.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans hypothetical protein

TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 1009 bp

Poly A stretch at pos. 990, polyadenylation signal at pos. 969

```

1 TGGGCCTGTG GCTGGGGGCA GAGCTCAGAC TGTCTTCTGA AGATTGATGT
51 CTATTTCCTT GAGCTCTTTA ATTTTGTTCG CAATTTGGAT AAACATGGCA
101 CAAATCCAGC AGGGAGGTCC AGATGAAAAA GAAAAGACTA CCGCACTGAA
151 AGATTTATTA TCTAGGATAG ATTTGGATGA ACTAATGAAA AAAGATGAAC
201 CGCCTCTTGA TTTTCCTGAT ACCCTGGAAG GATTTGAATA TGCTTTTAAT
251 GAAAAGGGAC AGTTAAGACA CATAAAACT GGGGAACCAT TTGTTTTTAA
301 CTACCGGGAA GATTTACACA GATGGAACCA GAAAAGATAC GAGGCTCTAG
351 GAGAGATCAT CACGAAGTAT GTATATGAGC TCCTGGAAAA GGATTGTAAT
401 TTGAAAAAAG TATCTATTCC AGTAGATGCC ACTGAGAGTG AACCAAAGAG
451 TTTTATCTTT ATGAGTGAGG ATGCTTTGAC AAATCCACAG AAATGATGG
501 TTTTAATTCA TGGTAGTGGT GTTGTGAGG CAGGGCAGTG GGCTAGAAGA
551 CTTATTATAA ATGAAGATCT GGACAGTGGC ACACAGATAC CGTTTATTAA
601 AAGAGCTGTG GCTGAAGGAT ATGGAGTAAT AGTACTAAAT CCAATGAAA
651 ACTATATTGA AGTAGAAAAG CCGAAGATAC ACGTACAGTC ATCATCTGAT
701 AGTTCAGATG AACCAGCAGA AAAACGGGAA AGAAAAGATA AAGTTTCTAA
751 AGTAACAAAG AAGCGACGTG ATTTCTATGA GAAGTATCGT AACCCCAAAA
801 GAGAAAAAGA AATGATGCAA TTGTATATCA GAGTGAGTGA GATCACTACT
851 TTCCTTFACT ATTTTCTTTA CCTGTATAT ATTTTATTAT ATGTAGATTG
901 TTTTGTTTTT CTTCAAGAAT ATTAATTCTT TTATTTGTCA TCATTTATTT
951 CCCATGCTCG TCTACTTGGG TTAAATGGGT TTTTAAATTC AAAAAAAAAA
1001 AAAAAAAAAA

```

BLAST Results

Entry I89937 from database EMBL:
Sequence 11 from patent US 5723315.
Score = 1083, P = 2.2e-42, identities = 223/231

Entry I89938 from database EMBL:
Sequence 12 from patent US 5723315.
Score = 875, P = 7.4e-33, identities = 175/175

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 922 bp; peptide length: 292
Category: similarity to unknown protein
Classification: unset

1 MSISLSSLIL LPIWINMAQI QGGGPDEKEK TTALKDLLSR IDLDELMKKD

```

51 EPPLDFPDTL EGFYAFNEK GQLRHIKTGE PFVFNREDL HRWNQKRYEA
101 LGEIITKYVY ELLEKDCNLK KVSIPVDATE SEPKSFIFMS EDALTNPQKL
151 MVLINGSGVV RAGQWARRLI INEDLDSGTQ IPFIKRAVAE GYGIVLNPV
201 ENYIEVEKPK IHVQSSSDSS DEPAEKREKR DKVSKVTKKR RDFYEKYRNP
251 QREKEMMQLY IRVSEITTFL YFYLVLVYIL LYVDCFVFLQ EY

```

BLASTP hits

Entry S67436 from database PIR:
 hypothetical protein - fission yeast (*Schizosaccharomyces pombe*)
 Length = 266
 Score = 112 (39.4 bits), Expect = 0.00037, P = 0.00037
 Identities = 33/147 (22%), Positives = 69/147 (46%)

Entry CEY75B8A.12 from database TREMBLNEW:
 gene: "Y75B8A.31"; *Caenorhabditis elegans* cosmid Y75B8A
 Score = 327, P = 1.5e-29, identities = 72/140, positives = 93/140

Alert BLASTP hits for DKFZphfbr2_2d17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2d17, frame 2

Report for DKFZphfbr2_2d17.2

```

[LENGTH]      292
[MW]           34260.50
[pI]           5.50
[HOMOL]        TREMBLNEW:AF064782_1 product: "unknown"; Mus musculus clone pEN87 unknown mRNA,
partial cds. 1e-119
[KW]           SIGNAL PEPTIDE 19
[KW]           TRANSMEMBRANE 1
[KW]           LOW_COMPLEXITY 10.96 %

```

```

SEQ  MSISLSSLLILPIWINMAQIQGGPDEKEKTTALKDLLSRIDLDELMMKKDEPPLDFPDTL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccchhhhhhhchhhhhhhccccccccccccchhhhhhhhhhhhhhhchhhhhhhcccccccccc
MEM  .....

```

```

SEQ  EGFYAFNEKGQLRHIKTGEFVFNREDLHRWNQKRYEALGEIITKYVYELLEKDCNLK
SEG  .....
PRD  hhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....

```

```

SEQ  KVSIPVDATESEPKSFIFMSDALTNPQKLMVLINGSGVVVRAGQWARRLIINEDLDSGTQ
SEG  .....
PRD  eeeeeccccccccccccccccccccccccccccccccchhhhhcccccccccccccccccc
MEM  .....

```

```

SEQ  IPFIKRAVAEGYGIVLNPVNPENYIEVEKPKIHVQSSSDSSDEPAEKREKDKVSKVTKKR
SEG  .....
PRD  chhhhhhhccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhh
MEM  .....

```

```

SEQ  RDFYEKYRNPQREKEMMQLYIRVSEITTFLYFYLVLVYILLYVDCFVFLQ EY
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhccccchhhhhhhhhhhhhheeeehhhhhhhhhhhhhheeeeeecccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

(No Prosite data available for DKFZphfbr2_2d17.2)

(No Pfam data available for DKFZphfbr2_2d17.2)

DKFZphfbr2_2d20

group: brain derived

DKFZphfbr2_2d20 encodes a novel 197 amino acid protein with similarity to *Synechocystis* sp. P74594 hypothetical132.8 kD protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to *Synechocystis* sp. (PCC 6803)

complete cDNA, complete cds, EST hits
potential start at bp 67 matches kozak consensus ANCatgG

Sequenced by Qiagen

Locus: unknown

Insert length: 1787 bp
Poly A stretch at pos. 1768, polyadenylation signal at pos. 1743

```
1 TGGGGCGGCC GCGGCGGGAA CATGGAGGAG CTGCTGAGGC GCGAGCTGGG
51 CTGCAAGCTCT GTCAGGGCCA CGGGCCACTC GGGGGCGGGG TGCATCAGCC
101 AGGGCCGGAG CTACGACACG GATCAAGGAC GAGTGTTCGT GAAAGTGAAC
151 CCCAAGGCGG AGGCCAGAAG AATGTTTGAA GGTGAGATGG CAAGTTTAAC
201 TGCCATCCTG AAAACAAACA CGGTGAAAGT GCCCAAGCCC ATCAAGGTTT
251 TGGATGCCCC AGGCGGCGGG AGCGTGCTGG TGATGGAGCA CATGGACATG
301 AGGCATCTGA GCAGTCATGC TGCAAAGCCT GGAGCCCAGC TGGCCGATTT
351 ACACCTTGAT AACAAAGAAG TTGGAGAGAT GCGCCTGAAG GAGGCGGGCA
401 CAGTGTGGAG AGGAGGTGGG CAGGAGGAAC GGCCCTTTGT GGCCCGGTTT
451 GGATTTGACG TGGTGACGTG CTGTGGATAC CTCCCCAGG TGAATGACTG
501 GCAGGAGGAC TGGGTCGTGT TCTATGCCCC GCAGCGCATT CAGCCCCAGA
551 TGGACATGGT GGAGAAGGAG TCTGGGGACA GGGAGGCCCT CCAGCTTTGG
601 TCTGCTCTGC AGTAAAAGAT CCCTGACCTG TTCCGTGACC TGGAGATCAT
651 CCCAGCCTTA CTCCACGGGG ACCTCTGGGG TGGAAACGTA GCAGAGGATT
701 CCTCTGGGCC GGTGATTTTT GACCCAGCTT CTTTCTACGG CCACTCGGAA
751 TATGAGCTGG CAATAGCTGG CATGTTTGGG GGCTTTAGCA GCTCCTTTTA
801 CTCCGCTTAC CACGGCAAAA TCCCCAAGGC CCCAGGATTC GAGAAGCGCC
851 TTCAGTTGTA TCAGCTCTTT CACTACTTGA ACCACTGGAA TCATTTTGGA
901 TCGGGGTACA GAGGATCCTC CCTGAACATC ATGAGGAATC TGGTCAAGTG
951 AGCGGGCCTT ACTCTGGAAG GAGGTCTCAG AGGTTTCTCC ACAGTCTCTT
1001 TCTGGGCAAA TTCTTGTTTC TTCACATGCC GGACTAGCTT AAGACCAATG
1051 CAGTAGCTTA TTTCCAAGCC TTGCAAAGTA TATAATATCT AAGAGGAAAG
1101 GTTTTGTGAT CCCAGCGTTG TCCACTTTGT GGGGCTTTGT AGGTAGACGG
1151 AGCCACACTA CAGGCAGGGT ATGAGCAGAG GGATGTATGG AGTGTGGGCG
1201 ACTCTGAGCC TCACTGCTGC TGCAAGGTGG GGAAACTGTA AGTGAACCCC
1251 TGTGGGTGCG GGGGAGGGTA TCCGGTGCGC AGGGAGGTGG CCAGCGCCCC
1301 CGGGCACCTG TGCTCATAGG TACCTTTCCG CTGCCTCCTC CTTGCTCTCC
1351 TGTGCAGGAA TGTCTCTGAG CTGTTTACGT TGATGCTTCT TGGTTGGCAA
1401 GACTTGGGTG TAGACATGAA ACCACCTTAC TAAAAGCGTC TAAAATGAC
1451 CAATTCCAGA ATCAAGCGTA TTCCGTTTTT CTCTGCATG ATCCCTGGGC
1501 CCTCCCGCAG GCTGAGCAAG TCTGTAAACT GATTCTGGGA GAAACCAAGC
1551 TGCTGGCCGT AGGATGTCCT TGGGTACATC CAGGAGTCTT CATTGCTTCT
1601 GTTATTACCC CGTCTCCTCT GCCATTTTCT ACAGCTTGCT GAGTTGTCAT
1651 TCCTTTGCAA CATTAAAATA CATGCTGAAC TCATATTTT CTTCCTTCA
1701 CTGTTGTAGT AAAGAGACAT ATTTTCATGAA TGGCATTGAT GCTAATAAAC
1751 CCTTTGCCCA AAAAATTGAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 22 bp to 612 bp; peptide length: 197
 Category: similarity to unknown protein
 Prosite motifs: LEUCINE_ZIPPER (117-139)

1 MEELLRRELG CSSVRATGHS GGCISQGRS YTDQGRV FV KVNPKAEARR
 51 MFE GEMASLT AILKTNTVKV PKPIKVLDAP GGGSVLVMEH MDMRHLSSHA
 101 AKLGAQLADL HLDNKKLGEM RLKEAGTVWR GGGQEERPFV ARFGFDVVTC
 151 CGYLPQVNDW QEDWVVFYAR QRIQPQMDMV EKESGDREAL QLWSALQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2d20, frame 1

Report for DKFZphfbr2_2d20.1

[LENGTH] 197
 [MW] 21963.25
 [pI] 6.96
 [HOMOL] PIR:S76790 hypothetical protein - Synechocystis sp. (strain PCC 6803) 9e-12
 [SUPFAM] hypothetical protein b1725 1e-06
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MYRISTYL 2
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 2
 [KW] Alpha_Beta

SEQ MEELLRRELGCSSVRATGHSGGCISQGRSYTDQGRV FV KVNPKAEARRMFE GEMASLT
 PRD ccchhhhhccccceeeccccceeeccccccccceeeccchhhhhhhhhhhhhhh

SEQ AILKTNTVKV PKPIKVLDAPGGGSVLVMEHMDMRHLSSHA AKLGAQLADLHLDNKKLGEM
 PRD hhhhhheeeccccceeeccccceeeccccccccchhhhhhhhhhhhhhhccccchhh

SEQ RLKEAGTVWRGGGQEERPFVARFGFDVVTC CGYLPQVNDWQEDWVVFYARQRIQPQMDMV
 PRD hhhhccccccccccccceeeccccceeeccccccccccccchhhhhhhhhhhhhhhhh

SEQ EKESGDREALQLWSALQ
 PRD hhhccchhhhhhhhhccc

Prosite for DKFZphfbr2_2d20.1

PS00002	20->24	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	13->16	PKC_PHOSPHO_SITE	PDOC00005
PS00005	67->70	PKC_PHOSPHO_SITE	PDOC00005
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00029	96->118	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_2d20.1)

DKFZphfbr2_2g18

group: brain derived

DKFZphfbr2_2g18 encodes a novel 229 amino acid protein with partial similarity to the humane dJ30M3.2 gene product.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

J30M3.2 extension of genmodel

complete cDNA, complete cds, EST hits
(mouse ESTs with >90% Identities)

Sequenced by Qiagen

Locus: /map="6p22.1-22"

Insert length: 2444 bp

Poly A stretch at pos. 2425, no polyadenylation signal found

```
1 TGGTCGAGGG TCGACGGTAT CGATAAGTTT TTTTTTTTTT TTTTTTTTTT
51 TGGAAAGCAA GGATCACA CTCCCTCC TGTTCCTTAA TCCCTTTTCT
101 AAAAAGGGGG GAAATCCGG ATGGATTTTA GGGATTGGTC TGGTGTACAG
151 TGTGTCTTAT TGCACACCTA AATCCTGATT ATAGGCTTTT CATTTCTCCG
201 CAAAGCCTTT ATTTTGGCAG TTAAGCCAAA TGTGTTTTCC AGAAAGTTAG
251 TATTTTCTC CTCTTTCTT CCTTCTTTC CTCCCTTTT CCCGTCTGAC
301 CCCAAACGTT ATGTGCCAAA CATGACTGGA CAGCAGCTTT TGTTCCTTGA
351 CCTGTGAATA TGACAGTCTG CTAATATTGA CAGAAGGTGC AGTTTTTGGG
401 TTATAGTCGT GATTTTCGCT AATCAATCAT ATTAGCAGGA AAAAAATGA
451 CTGTTTCTG TTGTACTTGA GTCCTAAGAA AAAGTGCCCA TAGTTTAGTG
501 ACAATTTCCA AAGGCTTTAG TACCACTGTG ATTTCAAAAT GGGGGACCCA
551 AACTCCCGGA AGAAACAAGC TCTGAACAGA CTACGTGCTC AGCTTAGAAA
601 GAAAAAAGAA TCTCTAGCTG ACCAGTTTGA CTTCAAGATG TATATTGCCT
651 TTGTATTCAA GGAGAAGAAG AAAAAGTCAG CACTTTTTGA AGTGTCTGAG
701 GTTATACCAG TCATGACAAA TAATTATGAA GAAATATCC TGAAGGTGT
751 GCGAGATTCC AGCTATTCTT TGGAAAGTTC CCTAGAGCTT TTACAGAAGG
801 ATGTGGTACA GCTCCATGCT CCTCGATATC AGTCTATGAG AAGGGATGTA
851 ATTGGCTGTA CTCAGGAGAT GGATTTTCTT CTTTGGCCTC GGAATGATAT
901 TGAATAAATC GTCTGTCTCC TGTTTTCTAG GTGGAAAGAA TCTGATGAGC
951 CTTTATAGCC TGTTCAGGCC AAATTTGAGT TTCATCATGG TGACTATGAA
1001 AAACAGTTTC TGCATGTACT GAGCCGCAAG GACAAGACTG GAATCGTTGT
1051 CAACAATCCT AACCACTCAG TGTTCCTCTT CATTGACAGA CAGCACTTGC
1101 AGACTCCAAA AAACAAAGCT ACAATCTTCA AGTTATGCAG CATCTGCCTC
1151 TACCTGCCAC AGGAACAGCT CACCCACTGG CGAGTTGGCA CCATAGAGGA
1201 TCACCTCCGT CCTTATATGC CAGAGTAGAG TACTGACCAG CAAAATGGAG
1251 AAGATCAGAG AATGCAGCAG CAGTTTTTTT TCTGTGTTTC TTACCACTTT
1301 ATTCTTTTCA AGTTTAAAGA AAATGGACTC ATGCACAGAA CACTATGCAT
1351 TTTGAAACTT GTTCATCTCG GATTTTTTTA AATCATTTTT ATCTCAGAAC
1401 TTAACAACAA ATTAGATGTC GTGCACGGAC TGTGTGAAAG AAGATGCTTT
1451 GCATATTGTC TGCATGTCAT CAGTATCTTA CTAAAAATGT GAAATGAAAG
1501 GACTATTGTA CACTGAAATG CTTAAATGTA TCTGAAAGCA CAAGGTGATA
1551 CTCATTTTTA TGGTCTTCCC ATTTGTGCTG GTTTTTGCCT CTTTGACATC
1601 TGTCATCAGT ATTTAGAGGG TGAGAAAGTA ATGTAACAGG TATAAATAAC
1651 ATTTTAAAAA ACAATAACTT TGCTATAATC ACAGTTGTTC CAGAGCACTG
1701 TCAGATACAT TCTAATGACC AGAAGTGGTT TAAAAAAGA AAATACAACC
1751 ATGGGAAAGA AATCTTAAAT GAAAAACGCA TCTCATTTGA GGCATTTTTG
1801 CCTCATATTT TACTGGGCCA TGTGTTGTTT CTGGTACTCA TGTATTTTTT
1851 TTTTTCAG ATCTCTTCC CCAAGTTGCT ATTGTAAGAG TATCTGCTG
1901 CGTGTGGATG CAGTTATACA CATTAAGCA GATCTGGAGT CTGAAGTAGC
1951 TATAAAGCAG CTATAAACA GAAATACATG CATAGCTGCA GAAACCATGA
2001 TAGGTAGAGG ACTTTCTTT TGGTTTGTG TTTTGTGTT TTTTGTGTT
2051 TTTGGTTTTA CAGAGAAGAG ATTTTATTA CAAAGAAAAA AATTCCAGTG
2101 AATTGTGCAG AAATGCTGGT TTTTACACCA TCCTAAAGAA AAATTTTACA
2151 AGGGTGTGTT GGAGTAGAAA AAAGTTTATA AAGTTGGAAT CTTAAATTGT
2201 AAAATTAACC ATTGAGTGTC AAAGTTCTAA AAGCAGAACT CATTTCTGTC
2251 AATGAACATA AGGAAAGACT ACTGTATAGG TTTTTTTTTT TCTCCTTTTA
2301 AATGAAGAAA AGCTTTGCTT AAGGGTTGCA TACTTTTATT GGAGTAAATC
2351 TGAATGATCC TACTCCTTTG GAGTAAGACT AGTGCTTACC AGTTTCCAAT
2401 TGTATTAGC TTCTGTTGGA ATTTGAAAAA AAAAAAATAA AAAA
```

BLAST Results

Entry HS338352 from database EMBL:
human STS EST171398.
Score = 1747, P = 3.0e-74, identities = 359/365

Entry HS447255 from database EMBL:
human STS SHGC-10143.
Score = 1717, P = 6.5e-73, identities = 365/383

Entry HS30M3 from database EMBLNEW:
Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands.
Score = 6646, P = 0.0e+00, identities = 1344/1355

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 539 bp to 1225 bp; peptide length: 229
Category: putative protein

```

1 MGDPNRKKQ ALNRLRAQLR KKKESLADQF DFKMYIAFVF KEKKKKSALF
51 EVSEVIPVMT NNYEENILKG VRDSSYSLES SLELLQKDVV QLHAPRYQSM
101 RRDVIGCTQE MDFILWPRND IEKIVCLLFS RWKESDEPFR PVQAKFEFHH
151 GDYEKQFLHV LSRKDKTGIV VNNPNQSVFL FIDRQHLQTP KNKATIFKLC
201 SICLYLPQEQ LTHWAVGTIE DHLRPYMPE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2g18, frame 2

TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands., N = 1, Score = 470, P = 1.1e-44

>TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands.
Length = 86

HSPs:

Score = 470 (70.5 bits), Expect = 1.1e-44, P = 1.1e-44
Identities = 86/86 (100%), Positives = 86/86 (100%)

```

Query: 144 AKFEFHHGDYEKQFLHVLRSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC 203
        AKFEFHHGDYEKQFLHVLRSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC
Sbjct: 1 AKFEFHHGDYEKQFLHVLRSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC 60

Query: 204 LYLPQEQLTHWAVGTIEDHLRPYMPE 229
        LYLPQEQLTHWAVGTIEDHLRPYMPE
Sbjct: 61 LYLPQEQLTHWAVGTIEDHLRPYMPE 86

```

Pedant information for DKFZphfbr2_2g18, frame 2

Report for DKFZphfbr2_2g18.2

```
SEQ      MGDPNRKKQALNRLRAQLRKKKESLADQDFDKMYIAFVFKEKKKSALFEVSEVIPVMT
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhheeeec

SEQ      NNYEENILKGVRDSSYSLESSLELLQKDQVVLHAPRYQSMRRDVIGCTQEMDFILWPRND
SEG      .....XXXXXXXXXX.....
PRD      cchhhhhhhcccccchhhhhhhhhhhhhhhhhhhcccccceeeccccceeecccch

SEQ      IEKIVCLLSRWKESDEPFRPVQAKFEFHGGDYEQFLHLVLSRKDKTGIVNNPNQSVFL
SEG      .....
PRD      hhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhcccccceeeccccceeee

SEQ      FIDRQHLLQTPKNKATIFKLCSICLYLPQEQLTHWAVGTIEDHLRPYMPF
SEG      .....
PRD      eeeccccccccccccceeeeeeeeeeeccccccccceeecccccccccc
```

PS000001	175->179	ASN_GLYCOSYLATION	PDOC000001
PS000004	22->26	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	44->48	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	6->9	PKC_PHOSPHO_SITE	PDOC000005
PS000005	99->102	PKC_PHOSPHO_SITE	PDOC000005
PS000005	162->165	PKC_PHOSPHO_SITE	PDOC000005
PS000005	189->192	PKC_PHOSPHO_SITE	PDOC000005
PS000006	25->29	CK2_PHOSPHO_SITE	PDOC000006
PS000006	80->84	CK2_PHOSPHO_SITE	PDOC000006
PS000006	162->166	CK2_PHOSPHO_SITE	PDOC000006
PS000006	218->222	CK2_PHOSPHO_SITE	PDOC000006
PS000007	69->77	TYR_PHOSPHO_SITE	PDOC000007
PS000008	70->76	MYRISTYL	PDOC000008
PS000008	168->174	MYRISTYL	PDOC000008

227

DKF2phfbr2_2h1

group: brain derived

DKF2phfbr2_2h1 encodes a novel 180 amino acid protein with weak similarity to C.elegans D2007.4 protein

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C.elegans D2007.4 protein

CpG island in 5' region, complete cDNA

Sequenced by Qiagen

Locus: unknown

Insert length: 957 bp

Poly A stretch at pos. 939, polyadenylation signal at pos. 916

```

1 GGGGGTCCCT GACTTTATAT GGCTGCTCCT GCGGAGCGAC TGAGTCGTCC
51 GTGAGGAAAA AGAGGCGAGG CTTTCCGAG ATCGTCTCAG CGATGGCGCT
101 TCGGTCGCGG TTTTGGGGGT TGTTCCTCGG TTGCAGGAAC CCTGGGTGCA
151 GGTTCGCAGC CCTGTCAACC AGCTCCGAGC CGGCAGCGAA ACCTGAAGTG
201 GACCTGTGG AAAATGAAGC TGTCGCCCCA GAATTCACCA ACCGGAACCC
251 CCGGAACCTG GAGCTTTTGT CTGTAGCCAG GAAAGAGCGG GGCTGGCGGA
301 CGGTGTTTCC CTCCCGTGAG TTCTGGCACA GGTTCGAGT TATAAGGACT
351 CAGCATCATG TAGAAGCACT TGTGGAGCAT CAGAAATGGCA AGGTTGTGGT
401 TTCGGCCTCC ACTCGTGAGT GGGCTATTAA AAAGCACCTT TATAGTACCA
451 GAAATGTGGT GGCTTGTGAG AGTATAGGAC GAGTGCTGGC ACAGAGATGC
501 TTAGAGGCGG GAATCAACTT CATGGTCTAC CAACCAACCC CGTGGGAGGC
551 AGCCTCAGAC TCGATGAAAC GACTACAAAG TGCCATGACA GAAGGTGGTG
601 TGGTTCTACG GGAACCTCAG AGAATCTATG AATAAATGGA AGCATTAAAT
651 GTTTTGAACA TGTAATATA AATCTGTCAG CCACTACAGC CATCAAAAGA
701 GAGCATCTGG AAGAACAGCC AGCTTGGAAG TTTTACAGCA ATAATGTTGC
751 AGTGGAAATAT TATTTGTAGT TAAGGTCATC CTCCTCCCCT TTCTGTTTTT
801 TTAAATCAAG AACTACGTTT TGCCCTCTCT TTGGGCTTCA GAAGCATCTA
851 AGAAAAGCAG TCATCAATTA TAATTAACCT TCAAAGGGCA AGTCAGAAGT
901 TGTTTATAAA TTACAAAATA AAGGCATATT ATGAACCTTA AAAAAAAAAA
951 AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 93 bp to 632 bp; peptide length: 180
 Category: similarity to known protein
 Classification: unset

```

1 MALRSRFWGL FSVCRNPGCR FAALSTSSEP AAKPEVDPVE NEAVAPEFTN
51 RNPRNLELLS VARKERGWRT VFPSREFWHR LRVIRTQHHV EALVEHQNGK
101 VVVSASTREW AIKKHLYSTR NVVACESIGR VLAQRCLLEAG INFMVYQPTP
151 WEAASDSMKR LQSAMTEGGV VLREPQRIYE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_2h1, frame 3

PIR:S44789 D2007.4 protein - *Caenorhabditis elegans*, N = 1, Score = 194, P = 2e-15

PIR:JC5753 ribosomal protein L18 - *Vibrio proteolyticus*, N = 1, Score = 121, P = 1.1e-07

>PIR:S44789 D2007.4 protein - *Caenorhabditis elegans*
Length = 170

HSPs:

Score = 194 (29.1 bits), Expect = 2.0e-15, P = 2.0e-15
Identities = 51/134 (38%), Positives = 78/134 (58%)

```
Query:   48 FTNRNPRNLELLSVARKERGWRVTFP--SREFWHRLRVIRTQHHVEA-LVEHQNGKVVVS 104
          F NRNPRN EL+      G++      +R + +++ ++ + H E LV +Q+G VV+S
Sbjct:   9 FVNRNPRNNELMGRQAPNTGYQFEKDRAARSYIYKVELVEGKSHREGRLVHYQDG-VVIS 67

Query:  105 ASTREWAIAKKHLYSTRNVVACESIGRVLAQRCL EAGINFMVYQPTPWEAASDSMKRLQ-- 162
          AST+E +I  LYS + A +IGRVLA RCL++GI+F + T EA S +
Sbjct:  68 ASTKEPSIASQLYSKTDTSAALNIGRVLALRCLQSGIH FAMPGATK-EAIEKSQHQTHFF 126

Query:  163 SAMTEGGVVLRREPQRI 178
          A+ E G+ L+EP +
Sbjct: 127 KALEEEGLTLKEPAHV 142
```

Pedant information for DKFZphfbr2_2h1, frame 3

Report for DKFZphfbr2_2h1.3

```
[LENGTH]      180
[MW]           20576.57
[pI]           9.63
[HOMOL]        PIR:S44789 D2007.4 protein - Caenorhabditis elegans 2e-13
[FUNCAT]       j mrna translation and ribosome biogenesis [H. influenzae, HI0794] 2e-04
[SUPFAM]       Escherichia coli ribosomal protein L18 8e-06
[KW]           Alpha_Beta
```

```
SEQ      MALRSRFGWGLFSVCRNPGCRFAALSTSSEPAAKPEVDPVENEAVAPEFTNRNPRNLELLS
PRD      cccccceeeeeeeeeccccccccccccccccccccccccccccccccccccccccchhhh

SEQ      VARKERGWRVTFPSREFWHRLRVIRTQHHVEALVEHQNGKVVVSASTREWAIAKKHLYSTR
PRD      hhhccccccccchhhhhhhhhccccchhhhhhhhhccccceeeechhhhhhhhhhhcc

SEQ      NVVACESIGRVLAQRCL EAGINFMVYQPTPWEAASDSMKRLQSAMTEGGVVLRREPQRIYE
PRD      ccceehhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhccceeecccccccc
```

(No Prosite data available for DKFZphfbr2_2h1.3)

(No Pfam data available for DKFZphfbr2_2h1.3)

DKFZphfbr2_2h10

group: brain derived

DKFZphfbr2_2h10 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2176 bp

Poly A stretch at pos. 2161, polyadenylation signal at pos. 2143

```

1  TGGGGAGTAT  TCTAATTATA  TTTTATATTT  AATAAATTAT  TTTTCTATTT
51  CTTTGTATATA  TTAAGTTGCA  CACTTGTTC  TTTTATCCAG  AAAGTTTAGT
101  ATAATAAAAA  TAGTTTAAAG  ATTAACGTG  AATGTAAAGG  AAAAGTATTA
151  TTAATTATTT  CAGGAAATTG  CAAGACCTAA  CATGGCTGAA  AGAGAAACAG
201  AAACATCAAA  TTCTGAAAGT  AAACAGATA  AAGCTGCTTC  TTCAAAAGAA
251  AAAAAATGGAT  GTAATGCAAA  TTCATTGAA  GGCTCATCAA  CAACAAAAG
301  TGAAGAAAGC  ATAACAGTTT  CAGATAAGGA  AAATGAAACC  TGTCTTGCAG
351  ACCAGGAAAC  TGGCTCAAAA  AACATCGTCA  GTTGTGATTC  AAATATTGGT
401  GCAGATAAAG  TGGAAAAGAA  AAAACAAATA  CAACACGTTT  GTCAGGAAAT
451  GGAGTTGAAG  ATGTGCCAGA  GTTCAGAAAA  CATAATCTTA  TCTGATCAGA
501  TTAAGATCA  CAACCTCCAGT  GAAGCCAGAT  TTTCTTCAAA  GAATATTAAG
551  GATTTGCGAT  TAGCATCAGA  TAATGTAAGC  ATTGATCAGT  TTTTGAGAAA
601  AAGACATGAA  CCTGAATCTG  TTAGTTCTGA  TGTAGCCGAG  CAAGGCAGTA
651  TTCATTGGA  ACCTCTGACT  CCATCCGAGG  TACTTGAGTA  TGAAGCCACA
701  GAGATTCTTC  AGAAAGGTAG  TGGTGATCCT  TCAGCCAAGA  CTGATGAAGT
751  AGTGTCTGAT  CAAACAGATG  ACATTCTCTG  AGGAAATAAC  CCTAGCACAA
801  CAGAGGCAAC  AGTAGACCTG  GAAGATGAAA  AAGAAAGAAG  TTGAAATTAG
851  TCATTTTAAG  TTTCACTGTA  CCAACGATAA  GGGCATTTGG  AACAGTGCTA
901  TCAGGTGAGC  TCAGTGGTGC  TGTGTAGGT  TCAGAAATGG  AAATATGTAA
951  GGGAGGTCAC  ACATACACTT  TACCTGTATG  TTCAACCTAT  GTTATCAAAC
1001  AAACCAATTC  ACCAATAATA  GCATGATTAG  TAGGGATTCC  CAAAAAGTTT
1051  TTAATAACAC  GAACAGGATT  TTAATGATAA  TTAATTTGCG  AGTGGAAAGG
1101  TCTCATTTAA  TGGTTTTCAA  GGAAATGGGA  TTTGGTTGCT  GACATGAATT
1151  GATGATATTA  GTAATATTTA  TAAAGCCTTT  CAAACTTCCA  TCAATCCTAA
1201  GCTAAAAAAT  TTTATTACCT  GTATATCCTT  TTCAGTTAAC  TGAGAGGAAG
1251  GGATTGGAAG  ACCATGTACT  TTTGGGGAGT  AATTGATTAA  AAACAATGGC
1301  TGATTGGCAT  TGTTAATGAA  GGCTTTATTT  GTGAGGATGA  TGCTGGTAAA
1351  TGGAGCATGC  TTAGAGTACT  AAATTGATCT  AATGAGAATT  TGGATGAACA
1401  TAAACTTAAT  TTTGGATTTA  ATATAACATT  CCAGTCAGAC  GCATGTAAAC
1451  AGAATATTTG  AATCTTTGTA  CCTCCATACA  AGTGTTAGCC  TGCCAGGCTG
1501  TAAGCTTACC  TTAATTAAAC  TTTCAGTGAA  AGTGGAAATTA  TTAAGATATA
1551  AATTTATATT  TGTGCTTTTT  GTCAGTGTGT  AAGCTGTGTA  GAAATTCCTT
1601  GATGTATTAG  TTGTATTAAT  GTAAAGTAGA  AATCCATTGT  TGAACCTCCT
1651  GTAGCTATTA  TGCTTTTAA  ATTGTTTTAA  TGTCTTCTCT  TAGAAATAGG
1701  CCCATAAAAA  TGGTCTGGAA  GCCAAACCAA  AGTATGGTAT  AATGTAGATA
1751  TTGTAAAGCA  GTAAACTGAA  AACATGTCCT  GGCATGTATT  CAGCCATGTT
1801  TAAGTGACTT  TTCTGTAATT  GTAAAATAAA  AACTTCAAAT  GGGACCTAAA
1851  ACAGTGATGT  AAAAGAAGCT  GTTTTGGAAA  TTTAGCCTAA  TTTATCTATA
1901  AGATGGCTGC  TAAATTGATT  TTTCAGTTCT  TTTTATCATC  TAAAAATATA
1951  TAGATATAGA  AATGAATAAT  ATGAAGAACA  GTAGTTTGCT  TTGAAATACT
2001  AATAAACTTT  TATTTAAGAT  GCTTCATTTT  TACTTCTTAA  AACGTGCTTT
2051  GGATTCTTAA  ATTTTGTTC  ACTGAATGTT  CAATGTTTTA  AATGGCGATT
2101  AAAATACTCT  GCTGTATATA  GTAGTTTTTG  AGTAAATATT  TGCAATAAAA
2151  ATCTGCCCCC  GAAAAAATA  AAAAA

```

BLAST Results

Entry G35287 from database EMBL:

human STS SHGC-37375.

Score = 2163, P = 2.8e-91, identities = 437/441

No Medline entry

Peptide information for frame 2

ORF from 182 bp to 841 bp; peptide length: 220
Category: putative protein

```

1 MAERETETSN SESKQDKAAS SKEKNGCNAN SFEGSSTTKS EESITVSDKE
51 NETCLADQET GSKNIVSCDS NIGADKVEKK KQIQHVQCEM ELKMCQSSSE
101 IILSDQIKDH NSSEARFSSK NIKDLRLASD NVSIQDFLRK RHEPEVSVD
151 VSEQSGSIHE PLTPSEVLEY EATEILQKGS GDPSAKTDEV VSDQTDIDPG
201 GNNPSTTEAT VDLEDEKERS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2 2h10, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 2h10, frame 2

Report for DKF2phfbr2 2h10.2

```
[LENGTH]      220
[MW]           24109.02
[pI]           4.51
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YKR092c] 4e-05
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKR092c] 4e-05
[PROSITE]      MYRISTYL 3
[PROSITE]      CK2_PHOSPHO_SITE 8
[PROSITE]      PKC_PHOSPHO_SITE 5
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         TNFR/NGFR cysteine-rich region
[KW]           Alpha Beta
```

SEQ MAERETETSNSESKQDKAASSKEKNGCNANSFEGSSTTKSEESITVSDKENETCLADQET
PRD cccccccccccccchhhhhhhhhcc

SEQ GSKNIVSCDSNIGADKVEKKQIQHVCQEMELKMCQSSENIILSDQIKDHNSSEARFSSK
PRD cccceeeccccchhhhhhhhhhhhhhhhhhhhcccceeecccccccccccccccccc

SEQ NIKDLRLASDNVSIQFLKRHEPESVSSDVSEQGSIHLEPLTPSEVLEYEATEILQKGS
PRD cchhhhhhhccchhhhhhhhhccccccccccccccccccccccccccccchhhhhhhcccccc

SEQ GDPSAKTDEVVSDQTDDI PGGNNPSTTEATVDLEDEKERS
PRD CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCeeehhhhhhccc

Prosites for DKF2phfbr2 2h10.2

PS000001	51->55	ASN_GLYCOSYLATION	PDOC000001
PS000001	111->115	ASN_GLYCOSYLATION	PDOC000001
PS000001	131->135	ASN_GLYCOSYLATION	PDOC000001
PS000005	20->23	PKC_PHOSPHO_SITE	PDOC000005
PS000005	37->40	PKC_PHOSPHO_SITE	PDOC000005
PS000005	47->50	PKC_PHOSPHO_SITE	PDOC000005
PS000005	118->121	PKC_PHOSPHO_SITE	PDOC000005
PS000005	184->187	PKC_PHOSPHO_SITE	PDOC000005
PS000006	9->13	CK2_PHOSPHO_SITE	PDOC000006
PS000006	13->17	CK2_PHOSPHO_SITE	PDOC000006
PS000006	20->24	CK2_PHOSPHO_SITE	PDOC000006
PS000006	38->42	CK2_PHOSPHO_SITE	PDOC000006
PS000006	45->49	CK2_PHOSPHO_SITE	PDOC000006
PS000006	47->51	CK2_PHOSPHO_SITE	PDOC000006
PS000006	163->167	CK2_PHOSPHO_SITE	PDOC000006
PS000006	205->209	CK2_PHOSPHO_SITE	PDOC000006
PS000008	26->32	MYRISTYL	PDOC000008

PS00008	34->40	MYRISTYL	PDOC00008
PS00008	201->207	MYRISTYL	PDOC00008

Pfam for DKFZphfbr2_2h10.2

HMM_NAME	TNFR/NGFR cysteine-rich region		
HMM	*CpeG.tYtD.WNHvpqClpCtrCePEMGQYmvqPCTwTQNTVC*		
	+E+ T +D +N ++C E G+ + +C+++ +		
Query	40	SEESITVSDKEN--ETC--LADQET--GSKNIVSCDSNIGADK	76

DKFZphfbr2_2i17

group: intracellular transport and trafficking

DKFZphfbr2_2i17.3 encodes a novel 201 amino acid putative GTP-binding protein related to Rab1B.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory(biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes. Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells.

The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

Medline

96245776: Intracellular transport and maturation of nascent low density lipoprotein receptor is blocked by mutation in the Ras-related GTP-binding protein, RAB1B

strong similarity to rab1

complete cDNA, complete cds, start at 47, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1985 bp

Poly A stretch at pos. 1901, polyadenylation signal at pos. 1859

```
1 GGGAGCAGAG TCGACTGGGA GCGACCGAGC GGGCCGCCGC CGCCGCCATG
51 AACCCCGAAT ATGACTACCT GTTTAAGCTG CTTTGATTG GCGACTCAGG
101 CGTGGGCAAG TCATGCCCTG TCCTGCCGTT TGCTGATGAC ACGTACACAG
151 AGAGCTACAT CAGCACCATC GGGGTGGACT TCAAGATCCG AACCATCGAG
201 CTGGATGGCA AAACATATCAA ACTTCAGATC TGGGACACAG CGGGCCAGGA
251 ACGGTTCCGG ACCATCACTT CCAGCTACTA CCGGGGGGCT CATGGCATCA
301 TCGTGGTGTA TGACGTCACT GACCAGGAAT CCTACGCCAA CGTGAAGCAG
351 TGGCTGCAGG AGATTGACCG CTATGCCAGC GAGAACGTCA ATAAGCTCCT
401 GGTGGGCAAC AAGAGCGACC TCACCACCAA GAAGGTGGTG GACAACACCA
451 CAGCCAAGGA GTTTGCAGAC TCTCTGGGCA TCCCCTTCTT GGAGACGAGC
501 GCCAAGAATG CCACCAATGT CGAGCAGGCG TTCATGACCA TGGCTGCTGA
551 AATCAAAAAG CGGATGGGGC CTGGAGCAGC CTCTGGGGGC GAGCGGCCCA
601 ATCTCAAGAT CGACAGCACC CCTGTAAAGC CGGCTGGCGG TGGCTGTTGC
651 TAGGAGGGGC ACATGGAGTG GGACAGGAGG GGGCACCTTC TCCAGATGAT
701 GTCCCTGGAG GGGGGAGGAG GTACCTCCCT CTCCCTCTCC TGGGGCATT
751 GAGTCTGTGG CTTTGGGGTG TCCTGGGCTC CCCATCTCCT TCTGGCCCAT
801 CTGCCTGCTG CCCTGAGCCC CGGTTCTGTC AGGGTCCCTA AGGGAGGACA
851 CTCAGGGCCT GTGGCCAGGC AGGGCGGAGG CCTGCTGTGC AGTTGCCTCT
901 AGGTGACTTT CCAAGATGCC CCCCTACACA CCTTCTTTTG GAACGAGGGC
951 TCTTCTGTCG GTGTCCCTCC CACCCCATG TATGCTGCAC TGGGTCTCT
1001 CCTTCTTCTT CCTGCTGTCC TGCCCAAGAA CTGAGGGTCT CCCCAGGCTC
1051 TACTGCCCTG GCTGCAGTCA GTGCCAGGG CGAGGAATGT GGCCAGGGGA
1101 TCCAGGACCT GGGATCCAGG GCCCTGGGCT GGACCTCAGG ACAGGCATGG
1151 AGGCCACAGG GGCCACAGC CCCACCTTT CCTCTCCCA CTGCCTCCTC
1201 TCCCTTCCCTA CACTCCAGC TCGAGCCGTC CAGCTGCGGT GGGATCTGAG
1251 TATATCTAGG GCGGGTGGGC GGGTAGCAGT GCTGGGCTG TGTCTTGAGC
1301 CTGGAGGGAG ACTGCTCCTG CCGCCCTCTG CCCTGCCGGA GACAGACCCA
1351 TGGCCTGCCT GCCCACCCTG CCCCTTTGTC CCCATGTGAG GCGGAGGCGG
1401 AAGGCCACAC GTGCCAGAGG CTGGGCACCA GCCTTAACCC TCACTCTGCT
1451 AGCACCTCCT CCCTTTCCCC AAGGTAGCAC ATCTGGCTCA CTCCTCACTC
1501 CGTCTCTGGA GCCCACCAGG GAAGGCCCTC ATCCCTGCTC GCTACTTCTC
1551 TGGGGAATGT GGGTTCCATC CAGGATTGGG GGCCTCTCTG CTCACCCACT
1601 CTGCACCCAG GATCCTAGTC CCCTGCCCTC TGGCACAGCT GCTTCCCTGA
1651 AGAAAGCAAG TCTTTGGTCT CCCTGAGAAG CCATGTCCCT CGTGCTGTCT
1701 CTTGCCTGTC CCACCTGTGC CCTGCCCTCC AGCTTGTATT TAAGTCCCTG
1751 GGCCTGCCCC TTGGGGTGCC CCCCGCTCCC AGGTTCCCTC CTGGTGTCTT
1801 GTCAGGCATT TTGCAAGGAA AAGCCACTTG GGGAAAGATG GAAAAGGACA
1851 AAAAAAATTA ATAAATTTC ATTGGCCCTC GGGTGAGCTG AGGGTTTTTG
1901 CAAGGAAAAA AAAAAA AAAA AAAA AAAA
1951 AAAAAA AAAAAGAAAA AAAAAA AAAA
```


BLAST Results

No BLAST result

Medline entries

91115900:
A family of ras-like GTP-binding proteins expressed in electromotor neurons.

Peptide information for frame 3

ORF from 48 bp to 650 bp: peptide length: 201
Category: strong similarity to known protein

```

1 MNPEYDYLFK LLLIGDSGVG KSCLLLRFD DTYTESYIST IGVD FKIRTI
51 ELDGKTIKIQ IWDTAGQERF RTITSSYYRG AHGIIVVYDV TDQESYANVK
101 QWLQEI DRYA SENVNKLLVG NKSDLTTKKV VDNTTAKEFA DSLGIPFLET
151 SAKNATNVEQ AFMTMAAEIK KRMGPGAASG GERPNLKIDS TPVKPAGGGC
201 C

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2i17, frame 3

SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B., N = 1, Score = 1023, P = 2.7e-103

PIR:S06147 GTP-binding protein rab1B - rat, N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:RAB1_DISOM RAS-RELATED PROTEIN ORAB-1., N = 1, Score = 967, P = 2.4e-97

PIR:TVHUYP GTP-binding protein Rab1 - human, N = 1, Score = 966, P = 3e-97

>SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B.
Length = 201

HSPs:

Score = 1023 (153.5 bits), Expect = 2.7e-103, P = 2.7e-103
Identities = 197/201 (98%), Positives = 199/201 (99%)

```

Query:      1 MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYTESYISTIGVD FKIRTI ELDGKTIKIQ 60
             MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYTESYISTIGVD FKIRTI ELDGKTIKIQ 60
Sbjct:      1 MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYTESYISTIGVD FKIRTI ELDGKTIKIQ 60

Query:      61 IWDTAGQERFRTITSSYYRG AHGIIVVYDVT DQESYANVKQWLQEI DRYASENVNKL LVG 120
             IWDTAGQERFRT+TSSYYRG AHGIIVVYDVT DQESYANVKQWLQEI DRYASENVNKL LVG 120
Sbjct:      61 IWDTAGQERFRTV TSSYYRG AHGIIVVYDVT DQESYANVKQWLQEI DRYASENVNKL LVG 120

Query:      121 NKSDLTTKKVVDNTTAKEFADSLGIPFLETS AKNATNVEQAFMTMAAEIKKRMGPGAASG 180
             NKSDLTTKKVVDNTTAKEFADSLG+PFLETS AKNATNVEQAFMTMAAEIKKRMGPGAASG 180
Sbjct:      121 NKSDLTTKKVVDNTTAKEFADSLGVPFLETS AKNATNVEQAFMTMAAEIKKRMGPGAASG 180

Query:      181 GERPNLKIDSTPVKPAGGGCC 201
             GERPNLKIDSTPVK A GGCC
Sbjct:      181 GERPNLKIDSTPVKSASGGCC 201

```

Pedant information for DKFZphfbr2_2i17, frame 3

Report for DKFZphfbr2_2i17.3

[LENGTH] 201

[MW] 22171.25
 [pI] 5.56
 [HOMOL] SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B. 1e-112
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL038c] 2e-77
 [FUNCAT] 30.08 organization of golgi [S. cerevisiae, YFL038c] 2e-77
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YER031c] 8e-46
 [FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YER031c] 8e-46
 [FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae, YGL210w] 1e-44
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YOR089c] 1e-30
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 9e-24
 [FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 9e-24
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-17
 [FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-17
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 1e-16
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR229c] 1e-11
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 1e-11
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 4e-10
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YNL180c] 9e-09
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YPL051w] 3e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL048c] 5e-05
 [BLOCKS] BL01019A ADP-ribosylation factors family proteins
 [BLOCKS] BL01115A GTP-binding nuclear protein ran proteins
 [SCOP] dlplk_ 3.25.1.3.1 cH-p21 Ras protein (human (Homo sapiens)) 2e-41
 [SCOP] dlguaa_ 3.25.1.3.10 Rap1A (Human (Homo sapiens)) 5e-60
 [SCOP] dlrrga_ 3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) (rat (Rattus)) 2e-30
 [SCOP] dlhura_ 3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) (human (Homo sapiens)) 2e-33
 [PIRKW] nucleus 1e-21
 [PIRKW] membrane trafficking 1e-110
 [PIRKW] oncogene 1e-25
 [PIRKW] endoplasmic reticulum 1e-105
 [PIRKW] phosphoprotein 1e-105
 [PIRKW] glycoprotein 3e-25
 [PIRKW] prenylated cysteine 1e-110
 [PIRKW] signal transduction 4e-23
 [PIRKW] transforming protein 1e-105
 [PIRKW] purine nucleotide binding 2e-24
 [PIRKW] alternative splicing 5e-26
 [PIRKW] P-loop 1e-110
 [PIRKW] lipoprotein 1e-110
 [PIRKW] proto-oncogene 3e-27
 [PIRKW] methylated carboxyl end 3e-27
 [PIRKW] hydrolase 7e-25
 [PIRKW] membrane protein 1e-105
 [PIRKW] GTP binding 1e-110
 [PIRKW] thiolester bond 5e-76
 [PIRKW] Golgi apparatus 1e-105
 [SUPFAM] ras transforming protein 1e-110
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 5
 [PROSITE] SIGMA54_INTERACT_1 1
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 3
 [PFAM] Ras_family (contains ATP/GTP binding P-loop)
 [KW] Alpha_Beta
 [KW] 3D

SEQ MNPEYDYLFKLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIELDGKTIKIQ
221p-EEEEEEETTTCHHHHHHHHHHHCCCCCCCCCTTTEEEE-EEEEETTEEEEEE

SEQ IWDTAGQERFRITITSSYYRGAHGIIVVYDVTQESYANVKQWLQEIDRYASENVNKLIVG
221p- EEECTTTTTTCGGGHHHHHHHCCCEEEEEETTBNHHHHHHHHHHHHHHHHHTTTTCEEEEE

SEQ NKSDLTTKKVVDNTTAKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG
221p- ETTTTCCTCC-CCCHHHHHHHHHHHCCCCCEEEETTTTTTHHHHHHHHHHHHHHH.....

SEQ GERPNLKIDSTPVKPAAGGCC
221p-

Prosites for DKFZphfbr2 2i17.3

PS00001	121->125	ASN_GLYCOSYLATION	PDOC00001
PS00001	133->137	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	56->59	PKC_PHOSPHO_SITE	PDOC00005
PS00005	126->129	PKC_PHOSPHO_SITE	PDOC00005
PS00005	135->138	PKC_PHOSPHO_SITE	PDOC00005
PS00005	151->154	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	179->183	CK2_PHOSPHO_SITE	PDOC00006
PS00007	27->34	TYR_PHOSPHO_SITE	PDOC00007
PS00008	18->24	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00017	15->23	ATP_GTP_A	PDOC00017
PS00675	11->25	SIGMA54_INTERACT 1	PDOC00579

Pfam for DKF2phfbr2 2i17.3

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM		*KLVLIGDSGVGKSCLLIRFTQNEFnEeYIPTIGvDFYtKTIEIDGktIK KL+LIGDSGVGKSCLL+RF +++++E+YI+TIGVDF++TIE+DGKTIK	
Query	10	LLLLIGDSGVGKSCLLLRFAADTTYTESYISTIGVDFKIRITTELDGKTIK	58
HMM		LQIWDTAGQERYRSMRPMYYRGAMGFLVYDITNRqSFENIrNwweEIrR LQIWDTAGQER+R++++YYRGA+G+++VVD+T+++S+ N+++W++EI+R	
Query	59	LQIWDTAGQERFRTITSSYYRGAHGIIVVYDVTDQESYANVKQLGQIDR	108
HMM		HCDrDENVPIMLVGNKCDLEDQRQVSTEEGQeFAREWGAIPFMETSAKTN +++ ENV ++LVGNK+DL +++V+ ++EFA+++G IPF+ETSAK++	
Query	109	YAS--ENVNKLVLGNKSDLTTKKVVDNTTAKEFADSLG-IPFLETSAKNA	155
HMM		iNVEEAFMEIvReIlqrMqe.q.NqteNinidQpsrnrk...rCCCIM* +NVE+AFM+++ EI++RM+ +++E +N++ +S++ K ++CC	
Query	156	TNVEAFMTMAAEIKKRMGPGAASGGERPNLKIDSTPVKPAAGGGC--	201

DKF2phfbr2_2k19

group: brain derived

DKF2phfbr2_2k19 encodes a novel 303 amino acid protein with similarity to human KIAA0378 product.

The protein contains a leucine zipper, which can mediate protein-protein-interaction. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to KIAA0378

encoded by the genomic clones HS147M19/HS608E8

Sequenced by Qiagen

Locus: unknown

Insert length: 1931 bp

Poly A stretch at pos. 1866, no polyadenylation signal found

```

1 GGGGGGGGCG CGCGGTGACA GCGCGGGGTT GCGGGCGTGG GACCCAGGGG
51 GCGACAGAGG CAGCAGCAGC CCGAGGCCTG AGGAGAGGAG ACCGGCGGCG
101 GCGGCAATGC TGGAGACCCT TCGCGAGCGG CTGCTGAGCG TGCAGCAGGA
151 TTTCACCTCC GGGCTGAAGA CTTTAAGTGA CAAGTCAAGA GAAGCAAAAG
201 TGAAAAGCAA ACCCAGGACT GTTCCATTTT TGCCAAAGTA CTCTGCTGGA
251 TTAGAATTAC TTAGCAGGTA TGAGGATACA TGGGCTGCAC TTCACAGAAG
301 AGCCAAAGAC TGTGCAAGTG CTGGAGAGCT GGTGGATAGC GAGGTGGTCA
351 TGCTTTCTGC GCACTGGGAG AAGAAAAAGA CAAGCCTCGT GGAGCTGCAA
401 GAGCAGCTCC AGCAGCTCCC AGCTTTAATC GCAGACTTAG AATCCATGAC
451 AGCAAATCTG ACTCATTAG AGGCGAGTTT TGAGGAGGTA GAGAACAACC
501 TGCTGCATCT GGAAGACTTA TGTGGGCAGT GTGAATTAGA AAGATGCAAA
551 CATATGCAGT CCCAGCAACT GGAGAATTAC AAGAAAAATA AGAGGAAGGA
601 ACTTGAAACC TTCAAAGCTG AACTAGATGC AGAGCACGCC CAGAAGGTCC
651 TGGAAATGGA GCACCCACAG CAAATGAAGC TGAAGGAGCG GCAGAAGTTT
701 TTTGAGGAAG CCTTCCAGCA GGACATGGAG CAGTACCTGT CCACTGGCTA
751 CCTGCAGATT GCAGAGCGGC GAGAGCCCAT AGGCAGCATG TCATCCATGG
801 AAGTGAACGT GGACATGCTG GAGCAGATGG TCCTGATGGA CATATCGGAC
851 CAGGAGGCCC TGGACGTCTT CCTGAACTCT GGAGGAGAAG AGAACACTGT
901 GCTGTCCCCC GCCTTAGGTA GGGTTGACAA ACTTGCAATTA GCTGAACCAG
951 GGCAGTATCG ATGCCACTCC CCTCCAAAGG TGAGACGTGA GAACCATCTG
1001 CCAGTCACTT ACGCATAAAC CCCCAAGCTC ACAGCCAGCT CCTGGCTCCC
1051 TAACCCACAG GTTCCACACG GCTGTGTGGC AGTGCAACA GTGGTGTGGT
1101 TCGTCATGTA ATTCTTCTCA AAGATTGAC ATGCTCCACT CCGGTAACCT
1151 TGGTGAGTTG AGAGCTTCT TGTGTGTTT CCTCCTTTA CCATCCAGAA
1201 ATCCATTGTA GTCTGCTCCT TGTGGTTAAG GACTGGCGTT TGCAGGGAGG
1251 TGCGGACTCT CCTGCGGGGC TCACGGGAAA CTCTTCCCTC TTCGTGCGAC
1301 AGGCATTAG GGGCGTGCC TCCATGGGCA AAGCCATGGT GTGTGTTTCA
1351 CTCTTGGCCT GTGTTGTAAA CTTAGTTGCA CTTCACTTCC TTTCATCCCT
1401 TCACAAAATT TTGTTTCACA TTCATGCAGC AAATATGGGC TGAGGTGCCA
1451 GACCTGTACC TGGGCTTGGT GCGTTTCAAA TTTCAGACCA GTTCTTTGGG
1501 CTGGGTCAAG GCAAAGCTCA GTCGTCCAG CAGCACCTCA GCCATCTGTA
1551 GAAGGTCTTA CCATTACCAC GGTTCAGCT TCCTCTAAAC TTCTCACCCG
1601 CTTCTCCTGG CAATCTGTCA GAACGGTGTC ATCTTGGGGA AGAGAAGGAG
1651 CTGGGTGCA TTTGCCCTCA TCCTGAGAAG GCCAGAATAC TGGAGACCAG
1701 CGTGAACCCT CACCCAGAGT CAGGGGAAGA TTTAGAAACA GTGACACCTG
1751 CATATAGAA TTTGATTCC TGAAGAGCCT ATTTAGTTCC ATAAAAATTG
1801 AGAAGTGTG AAGGTCAGTA ATTCCGACTT TCTCAGCAGT GGTGTCTCTG
1851 AATTACTGCA AAGGGTAAAA AAAAAAATA AAAAAACTTA TCGATACCGT
1901 CGACCTCGAT GATGATGATG ATGATGTGCA C

```

BLAST Results

Entry HS147M19 from database EMBL:
Homo sapiens DNA sequence from PAC 147M19 on chromosome 6p22.1-22.3.
Contains an unknown gene, ESTs and GSSs.
Score = 5540, P = 4.1e-275, identities = 1114/1120
3 exons 592-1884

Entry HS608E8 from database EMBL:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 608E8
Score = 797, P = 1.2e-78, identities = 161/163

6 exons 1-592

Medline entries

90294724:

The involucrin gene of the gibbon: The middle region shared by the hominoids

Peptide information for frame 2

ORF from 107 bp to 1015 bp; peptide length: 303

Category: similarity to known protein

Classification: unset

Prosites motifs: LEUCINE_ZIPPER (97-119)

```

1 MLETLRERLL SVQQDFTSGL KTLSDKSREA KVSKPRTVP FLPKYSAGLE
51 LLSRYEDTWA ALHRRRAKDC SAGELVDSEV VMLSAHWEKK KTSLEVELQEQ
101 LQQLPALIAD LESMTANLTH LEASFEEVEN NLLHLEDLCG QCELERCKHM
151 QSQQLENYKK NKRKELETFK AELDAEHAQK VLEMEHTQOM KLKERQKFFE
201 EAFQQDMEQY LSTGYLQIAE RREPIGSMSS MEVNVDMLEQ MVLMDISDQE
251 ALDVFNLSSG EENTVLSPAL GRVDKLALAE PGQYRCHSPP KVRRENHLPV
301 TYA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_2k19, frame 2

TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds., N = 1, Score = 137, P = 4.8e-06

PIR:I37037 involucrin - common gibbon, N = 1, Score = 124, P = 7.4e-05

PIR:A57013 early endosome antigen 1 - human, N = 1, Score = 128, P = 9.5e-05

>TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds.

Length = 808

HSPs:

Score = 137 (20.6 bits), Expect = 4.8e-06, P = 4.8e-06
Identities = 59/222 (26%), Positives = 103/222 (46%)

Query: 2 LETLRERLLSVQQDFTSGLKTL---SDKSREAKVKS-KPRTVPFLPKYSAGLELLSRYED 57
L TL E L S ++ LK D+ R +++S + K +A L+ E

Sbjct: 434 LATLEEAL-SEKERIIERLKEQREDDRELEEIESFRKENKDLKEKVNALQAELETES 492

Query: 58 TWAALHRRRAKDCASAGELVDSEVVMLSAHWEKKKTSLEVELQEQQLPALIADLESMTAN 117

+ L A ASAG DS++ L E+KK +L+ QL++ I D M
Sbjct: 493 SLIDLKEHASSLASAGLKRDSKLKSLAIEIQKKECKSLAQLKKAHN-IEDDSRMNPE 551

Query: 118 LTHLEASFEEVENNLLHLEDLCG--QCELERCKHMQSQQLENYKKNRK---ELETFAE 172

++++ + D CG Q E++R + +++EN K +K K ELE+
Sbjct: 552 FAD---QIKQLDKEASYRDECGKAQAEVDRLEIL-KEVENEKNDKDKKIAELESITLR 607

Query: 173 LDAEHAQKVLEMEHTQOMKLKERQKFFEEAFQQDMEQYLSSTGYLQIAE 220

+ +KV ++H QQ++ K+ + EE +++ ++ +LQI E
Sbjct: 608 HMKDQNKVANLKHNNQLEKKKNAQLLEEVRRREDSMADNSQHLQIEE 655

Score = 100 (15.0 bits), Expect = 6.2e-02, P = 6.0e-02
Identities = 44/156 (28%), Positives = 76/156 (48%)

Query: 57 DTWAALHRRRAKDCASAGELVDSEVVMLSAHWEKKKTSLEVELQEQQLPAL-IADLESMT 115

D A+ +R +C A VD + +L E +K + +L+ L + D
Sbjct: 560 DKEASYR--DECGKAQAEVDRLEILK-EVENEKNDKDKKIAELESITLRHMKDQNKKV 616

Query: 116 ANLTHLEASFEEVENNLLHLEDLCGQCE--LERCKHMQSQQLENYKKNRKKELETFKAEL 173

Pedant information for DKFZphfbr2 2k19, frame 2

Report for DKF2phfbr2_2k19.2

[illegible]

Prosites for DKFZphfbr2 2k19.2

PS00029 97->119 LEUCINE ZIPPER PDOC00029

(No Pfam data available for DKFZphfbr2 2k19.2)

DKFZphfbr2_2k14

group: cell cycle

DKFZphfbr2_2k14 encodes a novel 335 amino acid protein with strong similarity to rattus rattus IAG2 "implantation-associated protein" and the human N33 tumour-suppressor gene.

Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

The new protein can find application in modulating/blocking the cell cycle and in the therapy of tumours.

strong similarity to human N33 tumor suppressor gene

complete cDNA, complete cds, EST hits,
potential start at Bp 30 matches kozak consensus ANCatgG
potential transmembran protein (4 TM)
similarity to yeast OST3p (oligosaccharyltransferase gamma chain)

Sequenced by Qiagen

Locus: unknown

Insert length: 2241 bp

Poly A stretch at pos. 2221, no polyadenylation signal found

```

1 TGGGACTTAT AGAAGGGAGA GGAGCGAACA TGGCAGCGCG TTGGCGGTTT
51 TGGTGTGTCT CTGTGACCAT GGTGGTGGCG CTGCTCATCG TTTGCGACGT
101 TCCCTCAGCC TCTGCCCAAA GAAAGAAGGA GATGGTGTTA TCAGAAAAGG
151 TTAGTCAGCT GATGGAATGG ACTAACAAA GACCTGTAAT AAGAATGAAT
201 GGAGACAAGT TCCGTCGCCT TGTGAAAGCC CCACCGAGAA ATTACTCCGT
251 TATCGTCA TGCTGCTC TCCAACTGCA TAGACAGTGT GTCGTTTGCA
301 AGCAAGCTGA TGAAGAATTC CAGATCCTGG CAAACTCCTG GCGATACTCC
351 AGTGCATTCA CCAACAGGAT ATTTTGTGCC ATGGTGGATT TTGATGAAGG
401 CTCTGATGTA TTTCAGATGC TAAACATGAA TTCAGCTCCA ACTTTCATCA
451 ACTTTCTCTC AAAAGGGAAA CCCAAACGGG GTGATACATA TGAGTTACAG
501 GTGCGGGGTT TTTCAGCTGA GCAGATTGCC CGGTGGATCG CCGACAGAAC
551 TGATGTCAAT ATTAGAGTGA TTAGACCCCC AAATTATGCT GGTCCCTTAA
601 TGTGTTGGAT GCTTTTGGCT GTTATTGGTG GACTTGTGTA TCTTCGAAGA
651 AGTAATATGG AATTTCTCTT TAATAAAACT GGATGGGCTT TTGCAGCTTT
701 GTGTTTGTGT CTGCTATGA CATCTGGTCA AATGTGGAAC CATATAAGAG
751 GACCCACATA TGCCCATAG AATCCCCACA CGGGACATGT GAATTATATC
801 CATGGAAGCA GTCAAGCCCA GTTTGTAGCT GAAACACACA TTGTTCTTCT
851 GTTTAATGGT GGAGTTACCT TAGGAATGGT GCTTTTGTGT GAAGCTGCTA
901 CCTCTGACAT GGATATTGGA AAGCGAAAGA TAATGTGTGT GGCTGGTATT
951 GGACTTGTG TATTATCTT CAGTTGGATG CTCTCTATTT TTAGATCTAA
1001 ATATCATGGC TACCCATACA GCTTCTGAT GAGTTAAAAA GGTCCCAGAG
1051 ATATATAGAC ACTGGAGTAC TGGAAATTGA AAAACGAAAA TCGTGTGTGT
1101 TTGAAAAGAA GAATGCAACT TGTATATTCT GTATTACCTC TTTTTCACAA
1151 GTGATTTAAA TAGTTAATCA TTAAACCAAA GAAGATGTGT AGTGCCTTAA
1201 CAAGCAATCC TCTGTCAAAA TCTGAGGTAT TTGAAAATAA TTATCCTCTT
1251 AACCTTCTCT TCCCAGTGAA CTTTATGGAA CATTTAATTT AGTACAATTA
1301 AGTATATTAT AAAAATTGTA AACTACTAC TTTGTTTGTG TTAGAACAAA
1351 GCTCAAAACT ACTTTAGTTA ACTTGGTCAT CTGATCTTAT ATTGCCTTAT
1401 CCAAAGATGG GGAAGTAAG TCCTGACCAG GTGTTCCAC ATATGCCTGT
1451 TACAGATAAC TACATTAGGA ATTCATTCTT AGCTTCTTCA TCTTTGTGTG
1501 GATGTGTATA CTTTACGCAT CTTTCTTTT GAGTAGAGAA ATTATGTGTG
1551 TCATGTGGTC TTCTGAAAAT GGAACACCAT TCTTCAGAGC ACACGCTAG
1601 CCCTCAGCAA GACAGTTGTT TCTCCTCCTC CTTGCATATT TCCTACTGCG
1651 CTCCAGCCTG AGTGATAGAG TGAGACTCTG TCTCAAAAAA AAAGTATCTC
1701 TAAATACAGG ATTATAATTT CTGCTTGAGT ATGGTGTTAA CTACCTTGTA
1751 TTTAGAAAGA TTTCAGATTC ATTCCATCTC CTTAGTTTTC TTTTAAGGTG
1801 ACCCATCTGT GATAAAATA TAGCTTAGTG CTAATTCAG TGTAACTTAT
1851 ACATGGCCTA AAATGTTTCT ACAAATTAGA GTTTGTCAC TATTCCATTT
1901 GTACCTAAGA GAAAAATAGG CTCAGTTAGA AAAGGACTCC CTGGCCAGGC
1951 GCAGTGACTT ACGCCTGTAA TCTCAGCACT TTGGGAGGCC AAGGCAGGCA
2001 GATCAGAGG TCAGGAGTTC GAGACCATCC TGCCCAACAT GGTGAACCC
2051 CGTCTCTACT AAAAATATAA AAATTAGCTG GGTGTGGTGG CAGGAGCCTG
2101 TAATCCCAGC TGCACAGGAG GCTGAGGCAC GAGAATCACT TGAACCTCAG
2151 AGATGGAGGT TTCAGTGAGC CGAGATCACG CCACTGCAC TCAAGCCTGGC
2201 AACAGAGCGA GACTCCATCT CAAAAAATAA AAAAAAATAA A

```

BLAST Results

No BLAST result

Medline entries

96299740:
Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22.

97243398:
Tumour-suppressor genes in prostatic oncogenesis: a positional approach.

98334474:
Concordant methylation of the ER and N33 genes in glioblastoma multiforme.

Peptide information for frame 3

ORF from 30 bp to 1034 bp; peptide length: 335
Category: strong similarity to known protein

```

1 MAARWRFVCV SVTMVVALLI VCDVPSASAO RKKEMVLSEK VSQLEWNTNK
51 RPVIRMNGDK FRRLVKAPPR NYSVIVMFTA LQLHRQCVCV KQADEEFQIL
101 ANSWRYSSAF TNRIFFAMVD FDEGSDVFQM LNMNSAPTFI NFPAGKPKR
151 GDTYELQVRG FSAEQIARWI ADRTDVNIRV IRPPNYAGPL MLGLLAVIG
201 GLVYLRRSNM EFLFNKTGWA FAALCFVLAM TSGQMNHIR GPPYAHKNPH
251 TGHVNYIHGS SQAQFVAETH IVLLFNGGVT LGMVLLCEAA TSDMDIGKRK
301 IMCVAGIGLV VLFFSWMLSI FRSKYHGYPY SFLMS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_2k14, frame 3

TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein"; Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds., N = 1, Score = 1560, P = 3.4e-160

PIR:G02297 gene N33 protein - human, N = 1, Score = 1256, P = 5.6e-128

TREMBL:HSN33S11_1 gene: "N33"; product: "N33 protein form 2"; Human N33 protein form 2 (N33) gene, exon 11 and complete cds., N = 1, Score = 1252, P = 1.5e-127

>TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein"; Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds. Length = 308

HSPs:

Score = 1560 (234.1 bits), Expect = 3.4e-160, P = 3.4e-160
Identities = 295/307 (96%), Positives = 299/307 (97%)

```

Query:   29 AQRKKEMVLSEKVSQLEWNTNKRVPVIRMNGDKFRRLVKAPPRNYSVIVMFTALQLHRQCV 88
          AQRKKE VL EKV QLEWNTN+RPVIRMNGDKFR LVKAPPRNYSVIVMFTALQLHRQCV
Sbjct:   2  AQRKKEKVLVEKVIQLEWNTNQRVPVIRMNGDKFRPLVKAPPRNYSVIVMFTALQLHRQCV 61

Query:   89 VCKQADEEFQILANSWRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFPAGKGP 148
          VCKQADEEFQILAN WRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFP KGKP
Sbjct:   62 VCKQADEEFQILANFWRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFPKGP 121

Query:   149 KRGDYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS 208
          KR DTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS
Sbjct:   122 KRADTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS 181

Query:   209 NMEFLFNKTGWAFAALCFVLAMTSGQMNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE 268
          NMEFLFNKTGWAFAALCFVLAMTSGQMNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE

```


PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00008	193->199	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	278->284	MYRISTYL	PDOC00008
PS00009	296->300	AMIDATION	PDOC00009
PS00016	150->153	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2_2k14.3)

DKFZphfbr2_3c18

group: nucleic acid management

DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicase and RNA-dependent ATPase
from the DEAD box family

group helicases

Summary DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with similarity to DEAD-box subfamily ATP-dependent RNA helicases.
Deletion of the yeast homologue DBP5 is lethal.

strong similarity to RNA helicase and RNA-dependent ATPase from the
DEAD box family

complete cDNA, EST hits
complete cds ATG at Bp 109

Sequenced by AGOWA

Locus: /map="87.50 cR from top of Chr16 linkage group"

Insert length: 1713 bp
Poly A stretch at pos. 1696, no polyadenylation signal found

```
1 TGGGGTAGTG GGGCTGGAGC AGAGCCTGCC GCGAACCCCG GGAGCCCACG
51 ATCCCTTCGTG CCATCCCTCG AATCCACCAG CACGAGCGTC CCACCCGCGC
101 CTGGGACCAT GGCCACTGAC TCATGGGCCC TGGCGGTGGA CGAGCAGGAA
151 GCTGCGGCTG AGTCGTGAG CAACTTG CATTAAGGAAG AGAAAAATCAA
201 ACCAGATACC AATGGTGCTG TTGTCAAGAC CAATGCCAAT GCAGAGAAGA
251 CAGATGAAGA AGAGAAAGAG GACAGAGCTG CCCAGTCCTT ACTCAACAAG
301 CTGATCAGAA GCAACCTTGT TGATAACACA AACCAAGTGG AAGTCCTGCA
351 GCGGGATCCA AACTCCCTC TGTACTCGGT GAAGTCTTTT GAAGAGCTTC
401 GGTCCACACA GAACTTAATT GCCCAATCTC AGTCTGGTAC TGGTAAACAA
451 GCTGCCCTCG TGCTGGCCAT GCTTAGCCAA GTAGAACCCTG CAAACAAATA
501 CCCCCAGTGT CTATGTCTCT CCCCACGTA TGAGCTCGCC CTCCAAACAG
551 GAAAGTGAT TGAACAAATG GGCAAATTTT ACCCTGAAC TGAAGCTAGCT
601 TATGCTGTTT GAGGCAATAA ATTGGAAAGA GGCCAGAAGA TCAGTGAGCA
651 GATTGTGCTT GGCACCCCTG GGACTGTGCT GGACTGGTGC TCCAAGCTCA
701 AGTTTCATTG TCCCAAGAAA ATCAAGGTGT TTGTTCTGGA TGAGGCTGAT
751 GTCATGATAG CCACTCAGGG CCACCAAGAT CAGAGCATCC GCATCCAGAG
801 GATGCTGCCC AGGAAC TGCC AGATGCTGCT TTTCTCCGCC ACCTTTGAAG
851 ACTCTGTGTG GAAGTTTGCC CAGAAAAGTG TCCCAGACCC AAACGTTATC
901 AACTGAAGC GTGAGGAAGA GACCTGGAC ACCATCAAGC AGTACTATGT
951 CCTGTGCAGC AGCAGAGACG AGAAGTTCCA GGCCTTGCTT AACCTCTACG
1001 GGGCCATCAC CATTGCTCAA GCCATGATCT TCTGCCATAC TCGCAAAACA
1051 GCTAGTTGGC TGGCAGCAGA GCTCTCAAAA GAAGGCCACC AGGTGGCTCT
1101 GCTGAGTGGG GAGATGATGG TGGAAACAGG GGCTGCAGTG ATTGAGCGCT
1151 TCCGAGAGGG CAAAGAGAAG GTTTTGGTGA CCACCAACGT GTGTGCCCCG
1201 GGCATTGATG TTGAACAAGT GTCTGTCGTC ATCAACTTTG ATCTTCCCGT
1251 GGACAAGGAC GGGAAATCCTG ACAATGAGAC CTACCTGCAC CGGATCGGGC
1301 GCACGGGCCG CTTTGGCAAG AGGGGCTTGG CAGTGAACAT GGTGGACAGC
1351 AAGCAGACGA TGAACATCCT GAACAGAATC CAGGAGCATT TTAATAAGAA
1401 GATAGAAAGA TTGGACACAG ATGATTTGGA CGAGATTGAG AAAATAGCCA
1451 ACTGAGAAGC TCCACCAGCC ACTGATGCCA GCCCTGGCAC TGCCCCTGCA
1501 CAGGAGACAA GTGCGTTCAG GGCACAGGCC CCGACATCAC CCCAAGGACA
1551 ACGGCACAAG TAGAGAGAAA CTACCTACCT CACTTCAAAT TATGTTTGGG
1601 CTTGACAAAA ATGTATGCAA ATGATGGGGG ATGGTAGAAA AAAATTATTT
1651 ACACAACCTT GGAAGATTAG GCATGAATAC ACAGAGATTT ACCTTTAAAA
1701 AAAAAAAAAA AAA
```

BLAST Results

Entry G36496 from database EMBL:
 SHGC-53094 Human Homo sapiens STS cDNA.
 Length = 459
 Minus Strand HSPs:
 Score = 1693 (254.0 bits), Expect = 2.8e-70, P = 2.8e-70
 Identities = 369/387 (95%), Positives = 369/387 (95%)

Entry G44014 from database EMBLNEW:
 WIAF-3643-STS Human THudson SANGER Homo sapiens STS genomic, sequence
 tagged site.
 Score = 901, P = 2.3e-35, identities = 183/185

Medline entries

94192995:
 Gene 1994 Mar 25;140(2):171-177
 Mouse erythroid cells express multiple putative RNA helicase genes
 exhibiting
 high sequence conservation from yeast to mammals.

Peptide information for frame 1

ORF from 109 bp to 1452 bp; peptide length: 448
 Category: strong similarity to known protein

```

1 MATDSWALAV DEQEAAAESL SNLHLKEEKI KPDTNGAVVK TNANAEKTDE
51 EEKEDRAAQS LLNKLIRSNL VDNTNQVEVL QRDPNSPLYS VKSFEELRLP
101 QNLIAQSQSG TGKTAAFVLA MLQVEPANK YPQCLCLSPY YELALQTGKV
151 IEQMGKFYPE LKLAYAVRGN KLERGQKISE QIVIGTPGTV LDWCSKLKFI
201 DPKKIKVFVL DEADVMIATQ GHQDQSIRIQ RMLPRNCQML LFSATFEDSV
251 WKFAQKVVPD PNVIKLKREE ETLDTIKQYY VLCSSRDEKF QALCNLYGAI
301 TIAQAMIFCH TRKTASWLAA ELSKEGHQVA LLSGEMMVEQ RAAVIERFRE
351 GKEKVLVTN VCARGIDVEQ VSVVINFDLP VDKDGNPDNE TYLHRIGRTG
401 RFGKRGGLAVN MVDKSHSMNI LNRIQEHFNK KIERLDTDDL DEIEKIAN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3c18, frame 1

PIR:I49731 RNA helicase - mouse, N = 2, Score = 1758, P = 3.8e-223

TREMBL:AF005239_1 gene: "Dbp80"; product: "DEAD-box helicase";
 Drosophila melanogaster DEAD-box helicase (Dbp80) mRNA, complete cds.,
 N = 2, Score = 1142, P = 1.8e-125

SWISSPROT:YB66_SCHPO PUTATIVE ATP-DEPENDENT RNA HELICASE C12C2.06., N =
 2, Score = 911, P = 5.5e-103

PIR:S66920 probable RNA helicase CA5/6 - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 887, P = 1.9e-98

>PIR:I49731 RNA helicase - mouse
 Length = 478

HSPs:

Score = 1758 (263.8 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
 Identities = 338/349 (96%), Positives = 349/349 (100%)

```

Query: 100 PQNLIAQSQSGTGKTAAFVLAMLSQVEPANKYPQCLCLSPYELALQTGKVIEQMGKFYP 159
      PQNLIAQSQSGTGKTAAFVLAMLS+VEPA++YPQCLCLSPYELALQTGKVIEQMGKF+P
Sbjct: 130 PQNLIAQSQSGTGKTAAFVLAMLSRVEPADRYPQCLCLSPYELALQTGKVIEQMGKFHP 189

Query: 160 ELKLAYAVRGNKLERGQKISEQIVIGTPGTVDWCSKLKFDPPKKIKVFVLDEADVMIAT 219
      ELKLAYAVRGNKLERGQK+SEQIVIGTPGTVDWCSKLKFDPPKKIKVFVLDEADVMIAT
Sbjct: 190 ELKLAYAVRGNKLERGQKVFSEQIVIGTPGTVDWCSKLKFDPPKKIKVFVLDEADVMIAT 249

Query: 220 QGHQDQSIRIQRMPLPRNCQMLLFSATFEDSVWKFAQKVVPDPNVIKLKREETLDTIKQY 279

```

Sbjct: 250 QGHQDQSIRIQIR++PRNCQMLLFSAFEDSVWKFQKVVPDPN+IKLKREEETLDTIKQY
 QGHQDQSIRIQIRIVPRNCQMLLFSAFEDSVWKFQKVVPDPNIIKLKREEETLDTIKQY 309

Query: 280 YVLCSSRDEKFQALCNLYGAITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE 339
 YVLC++R+EKFQALCNLYGAITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE

Sbjct: 310 YVLCNNREEKFQALCNLYGAITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE 369

Query: 340 QRAAVIERFREGKEKVLVTTNVCARGIDVEQSVVINFDPVDKDGNDPNETYLHRIGRT 399
 QRAAVIERFREGKEKVLVTTNVCARGIDVEQSVVINFDPVDKDGNDPNETYLHRIGRT

Sbjct: 370 QRAAVIERFREGKEKVLVTTNVCARGIDVEQSVVINFDPVDKDGNDPNETYLHRIGRT 429

Query: 400 GRFGKRGGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN 448
 GRFGKRGGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN

Sbjct: 430 GRFGKRGGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN 478

Score = 419 (62.9 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
 Identities = 94/136 (69%), Positives = 104/136 (76%)

Query: 1 MATDSWALAVDEQEAASLSNLHLKEEKIKPDTNGAVVKTANAEKTDDEEKEDRAAQS 60
 MATDSWALAVDEQEAASLSNLHLKEEKIKPDTNGAVVKTANAEKTDDEEKEDRAAQS

Sbjct: 1 MATDSWALAVDEQEAASLSNLHLKEEKIKPDTNGAVVKTANAEKTDDEEKEDRAAQS 59

Query: 61 LLNKLIRSNLVDNTNQVEVLQDRPNSPLYSVKSFEELRL-PQNL---IAQSQSGTGKTAA 116
 LLNKLIRSNLVDNTNQVEVLQDRPNSPLYSVKSFEELRL PQ L A + K

Sbjct: 60 LLNKLIRSNLVDNTNQVEVLQDRPSSPLYSVKSFEELRLKPKQLLQGVYAMGFNRPSKIQE 119

Query: 117 FVLAMLSQVEPANKYPQ 133
 L M+ P N Q

Sbjct: 120 NALPMMLAEPQNLIAQ 136

Pedant information for DKFZphfbr2_3c18, frame 1

Report for DKFZphfbr2_3c18.1

[LENGTH] 448
 [MW] 50490.07
 [pI] 5.83
 [HOMOL] PIR:149731 RNA helicase - mouse 0.0
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YOR046c] 1e-102
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YDR021w] 2e-65
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR021w] 2e-65
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YJL138c] 1e-63
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YJL138c] 1e-63
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 2e-49
 [FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 9e-48
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YDL084w] 1e-43
 [FUNCAT] 1 genome replication, transcription, recombination and repair [H. influenzae, HI0892] 3e-39
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 1e-35
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 9e-27
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 8e-26
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 1e-23
 [FUNCAT] r general function prediction [M. jannaschii, MJ1401] 9e-08
 [FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 1e-05
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YMR190c] 1e-05
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YIR002c] 7e-04
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-64
 [PIRKW] RNA binding 1e-64
 [PIRKW] DEAD box 4e-64
 [PIRKW] transmembrane protein 3e-22
 [PIRKW] DNA binding 2e-32
 [PIRKW] ATP 1e-101
 [PIRKW] purine nucleotide binding 4e-64
 [PIRKW] P-loop 1e-101
 [PIRKW] hydrolase 4e-43
 [PIRKW] protein biosynthesis 1e-64
 [PIRKW] ATP binding 2e-35
 [SUPFAM] WW repeat homology 3e-29
 [SUPFAM] translation initiation factor eIF-4A 1e-64
 [SUPFAM] DEAD/H box helicase homology 1e-101
 [SUPFAM] DNA helicase recG 2e-06
 [SUPFAM] unassigned DEAD/H box helicases 1e-101
 [SUPFAM] ATP-dependent RNA helicase DBP1 9e-33

[SUPFAM] ATP-dependent RNA helicase DHH1 4e-48
 [SUPFAM] tobacco ATP-dependent RNA helicase DB10 3e-29
 [PROSITE] MYRISTYL 5
 [PROSITE] AMIDATION 1
 [PROSITE] CK2_PHOSPHO_SITE 6
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Helicases conserved C-terminal domain
 [PFAM] DEAD and DEAH box helicases
 [KW] Alpha_Beta

SEQ MATDSWALAVDEQEAAAESLSNLHLKEEKIKPDTNGAVVKTNANAECTDEEEKEDRAAQS
 PRD cccchhhhhhhhhhhhhhhccchhhhhhhccccceeeehhhhhhhhhhhhhhhhhhh

SEQ LLNKLIRSNLVDNTNQVEVLQRPNSPLYSVKSFEEELRLPQNLIQSQSGTGKTAFLVA
 PRD hhhhhhhhhccccceeeeeeccccceeeehhhhhhhhhccccceeeccccchhhhhhh

SEQ MLSQVEPANKYPQCLCLSPITYELALQTGKVEIQMGKFYPELKLAYAVRGNKLERGQKISE
 PRD hhhhhhhhhccccceeecccchhhhhhhhhhhhhhhccccceeeccccchhhhhhhhe

SEQ QIVIGTPTGTVLDWCSKLFIDPKKIKVFLDEADVMIATQGHQDQSIRIQRLPRNCQML
 PRD eeeccccchhhhhhhhhhhccccceeecccchhhhhhhccchhhhhhhhhccccceee

SEQ LFSATFEDSVWKFAQKVVPDPNVIKLKREEETLDTIKQYVLCSSRDEKFOALCNLYGAI
 PRD eeccccchhhhhhhhhhhccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ TIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTN
 PRD hhhhhheecchhhhhhhhhhhhhccccceeecccchhhhhhhhhhhccccceeeec

SEQ VCARGIDVEQSVVINFDPVDKDGNDNETYLHRIQRTGRFGKRGGLAVNMVDSKHSMMNI
 PRD cccccceeeeeeccccccccccccceeeeeeccccccccceeeeeecccchhhhh

SEQ LNRIQEHFNKKIERLDTDDLDEIEKIAN
 PRD hhhhhhhhhhhccccccccchhhhhccc

Prosites for DKFZphfbr2_3c18.1

PS00001	389->393	ASN_GLYCOSYLATION	PDOC00001
PS00002	109->113	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	226->229	PKC_PHOSPHO_SITE	PDOC00005
PS00005	275->278	PKC_PHOSPHO_SITE	PDOC00005
PS00005	284->287	PKC_PHOSPHO_SITE	PDOC00005
PS00005	311->314	PKC_PHOSPHO_SITE	PDOC00005
PS00005	399->402	PKC_PHOSPHO_SITE	PDOC00005
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	284->288	CK2_PHOSPHO_SITE	PDOC00006
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	175->181	MYRISTYL	PDOC00008
PS00008	185->191	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	406->412	MYRISTYL	PDOC00008
PS00009	402->406	AMIDATION	PDOC00009

Pfam for DKFZphfbr2_3c18.1

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLPpWILRnIyeMGFEkPTPIQQqAIPiILeG...RDVMACAQTGSGK		
	++ ++ +N ++	P	E+ +++A++Q+G+GK
Query	65	LIRSNLVDNTNQVEVLQRPNSPLYSVKSFEEELRLPQNLIQSQSGTGK	113
HMM	TAAFLIPMLQHIDwdPwqpPQdPrALILAPTRELAMQIEEcRkFgkHM		
	TAAF++ ML+++ +	+ PQ	+L L+PT ELA+Q+ +++++GK++
Query	114	TAAFLVLAQLSQVEPAN--KYPQ---CLCLSPITYELALQTGKVEIQMGKFY	158
HMM	ngIRImciYGGtnMRdQMRmLeRgPpHIVATPGRLLIDHIER.gtldLDr		
	+ + + + ++	++ +++ ++	+IVI+TPG ++D + +D ++

Query 159 PELKLAYAVR---GNKLERGQKISEQIVIGTPGTVLDWCSKLFIDPKK 204
HMM IeMLVMDEADRMLD.MGFIDQIRrIMrqIPMpwnRQTMMFSATMPdeIqE
I+++V+DEAD M+ +G +DQ RI R++P +N Q ++FSAT+ D++ +
Query 205 IKVFVLDEADVMIATQGHQDQSIRIQRMPL--RNCQMLLFSATFEDSVWK 252
HMM LARrFMRNPiRInIdMdElTtnEnIkQwYiyVerEMWKfdCLcrLlie*
+A ++ +P I ++++E T++ +IKQ+Y+ + + ++KF +LC+L++
Query 253 FAQKVVPDPNVIKLKREEETLD-TIKQYVVLCSRRDEKFQALCNLYG 298

HMM_NAME Helicases conserved C-terminal domain

HMM *EileeWLknLGirvmYIHGdMpQeERdeIMddFNnGEynVLicTDVggr
+L+ +L+++G +V+ + G M+ E+R ++++F++G+ +VL++T+V +R
Query 316 SWLAAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTNVCAR 364
HMM GIDIPdVNHVINYDM...PWNPEq..YIQRIGRTgRIG*
GID+++V++VIN+D+ + NP++ Y++RIGRTGR+G
Query 365 GIDVEQVSVVINFDLFPVDKDGNDNETYLHRIGRTGRFG 403

Medline

PMID: 10322435

"Unwinding RNA in : DEAD-box proteins and related families." de la Cruz J, Kressler D, Linder P

DKFZphfbr2_3f16

group: brain derived

DKFZphfbr2_3f16 encodes a novel 127 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1514 bp

Poly A stretch at pos. 1454, polyadenylation signal at pos. 1434

```
1 GGGGGGACTG GAGAAGGGAG GCGGCGGGCG AAGCGCACGT CGAGCGGGGG
51 AGCGGCGCTG CCTGTGGAGA TCCGCGGGAG CCGACAGGAT TCGTTGGCTG
101 CCGTCCCCGC TGCTGTGCAT TGGGTTAAAA ACGACAACCA ACATCAGCCA
151 TGAAAGATCC AAGTCGCAGC AGTACTAGCC CAAGCATCAT CAATGAAGAT
201 GTGATTATTA ACGGTCATTG TCATGAAGAT GACAATCCAT TTGCAGAGTA
251 CATGTGGATG GAAAAATGAAG AAGAAATCAA CAGACAAATA GAAGAGGAGT
301 TATGGGAAGA AGAATTATTT GAACGCTGTT TCCAAGAAAT GCTGGAAGAG
351 GAAGAAGAGC ATGAATGGTT TATTCCAGCT CGAGATCTCC CACAAACTAT
401 GGACCAAATC CAAGACCAGT TTAATGACCT TGTTATCAGT GAAGGCTCTT
451 CTCTGGGAAG TCTTGTGGTC AAGAGCAATC TGAATCCAAA TGCAAAGGAG
501 TTTGTTCCCT GGGTGAAGTA CGGAAATATT TGAGTAGACG GGGCCCTCTT
551 TTGGTGGATG TAGCACAATT TCCACACTGT GAAGGCAGTA TTAGAAGACT
601 TAATTGTAAA AGCACTCTTG TCACTGTGTT ACACTTATGC ATTGCCAAAG
651 TTTTGTAGTG TCTTGCATGC TTAATAAAG TGCTGAGACT GTTACTAAGT
701 AAAAAGCTGT CAACATTTA CTGAAATAG AATTGGCCCC ATGGCTTGAT
751 GTGAAGACAG CAAGGAAAGA AGCACCAGTC AAGTTGTGAA CAAGCACCAG
801 ATTAAGAGAC CTAACCTTTA CCAAAATGTC TTTTGTGAG GCTAATCTAT
851 CACTTGTAAA TGTCTAACT TTAATAATCAG TACATTAAAT TTGAGTTCCA
901 ACTGTTAAGC ATATTTCTCA GACTTAAAT TGATTATGTC CCCATCAAAA
951 AGAATCTCCA TTTCTGAAG GTCTGTAGT TAATTTGAGA TAATTTGTGA
1001 AAGGCAAGTA TGTCAATTA CTGAGGCTAC AAGTTAGTCA GCAGATGAGT
1051 GCCAGTCCAG CCTTTCCCGG TATGTTATTG TTAGAAATAT TGAGTTCTAA
1101 TGTTACATCT GAGGAAGTAT GTAATTTGAG AATTGTAAC TCTAAGGGAT
1151 TCACTGCATC ATAGCTATGC CTGTATGGAG TCTAACATAT GACCAATACC
1201 AACCATAAT CCAGCTGAAC AAAGATACTG TAACATTATG ATTTGAGTGG
1251 TGCTTTTCTT TGCTTTGTTA ACCATCAGCA GAGTCTGCAG CACAACCTTT
1301 AACAAAGCTA GAACAGTTT GGCTTCTTAA ACTTCATATT TGGGTAGGTT
1351 AAGCTGCCAT ACGTGTTCAG TGTGAATAGT GTTTAAGTTG AAAATATTGT
1401 AAAAAAATTA TATTTTTC AAAAAATTTA AAAAAATAAA TAATAGTAGA
1451 ACTGAAAAAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAGAAAAA
1501 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAGAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 150 bp to 530 bp; peptide length: 127
Category: putative protein

1 MKDPSRSSTS PSIINEDVII NGHSHEDDNP FAEYMMWENE EEFNRQIEEE

51 LWEEEFIERC FQEMLEEEEE HEWFIPARDL PQTMDQIQDQ FNDLVISEGS
101 SLEDLVVKS NLPNAKEFVP GVKYGNI

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3f16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_3f16, frame 3

Report for DKFZphfbr2_3f16.3

[LENGTH] 127
[MW] 14998.41
[pI] 4.04
[BLOCKS] BL01269D
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 2
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 27.56 %

SEQ MKDPSRSSTSPSIINEDVIINGSHEDDNPFAYEMWMENEEFNRIEEELWEEEFIERC
SEGXXXXXXXXXXXXXXXXXXXXX
PRD ccc

SEQ FQEMLEEEEEHEWFIPARDLPQTMDQIQDQFNDLVISEGSLEDLVVKS NLPNAKEFVP
SEG xxx
PRD hhhhhhhhhhhhhcc

SEQ GVKYGNI
SEG
PRD ccccccc

Prosite for DKFZphfbr2_3f16.3

PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	100->104	CK2_PHOSPHO_SITE	PDOC00006
PS00008	121->127	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_3f16.3)

DKFZphfbr2_3g8

group: metabolism

DKFZphfbr2_3g8.1 encodes a novel 178 amino acid protein with similarity to yeast ARD1 protein.

In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. The new protein could be part of this or an other NAT complex.

The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

strong similarity to N-TERMINAL ACETYLTTRANSFERASE COMPLEX ARD1 homolog

complete cDNA, complete cds? start at Bp 40, EST hits

Sequenced by AGOWA

Locus: /map="20"

Insert length: 1030 bp

Poly A stretch at pos. 1013, no polyadenylation signal found

```

1 TGGGCTTGGC GAACGGTCTT CGGAAGCGGC GGGGGCGCGA TGACCACGCT
51 ACGGGCCTTT ACCTGCGACG ACCTGTTCCG CTTCAACAAC ATTAACCTGG
101 ATCCACTTAC AGAAACTTAT GGGATTCCCT TCTACCTACA ATACCTCGCC
151 CACTGGCCAG AGTATTTCAT TGTTCAGTG GCACCTGGTG GAGAATTAAT
201 GGGTTATATT ATGGGTAAAG CAGAAGGCTC AGTAGCTAGG GAAGAATGGC
251 ACGGGCACGT CACAGCTCTG TCTGTTGCCC CAGAATTCG ACGCCTTGGT
301 TTGGCTGCTA AACTTATGGA GTTACTAGAG GAGATTTCAG AAAGAAAGGG
351 TGGGTTTTTT GTGGATCTCT TTGTAAGAGT ATCTAACCAA GTTGCAGTTA
401 ACATGTACAA GCAGTTGGGC TACAGTGTAT ATAGGACGGT CATAGAGTAC
451 TATTCGGCCA GCAACGGGGA GCCTGATGAG GACGCTTATG ATATGAGGAA
501 AGCACTTTC AGGGATACTG AGAAGAAATC CATCATACCA TTACCTCATC
551 CTGTGAGGCC TGAAGACATT GAATAACCC TGGCAGTGGT TCTTAGGCAG
601 ATACTCTAGA TGCTTTATGG ACAATATTAT TTTCATTGGA TGATTCTGGA
651 GCTCTATTAG GAGAAAAGTA ATCATTTTAG GTCTTAAAGA CTTCAAGAAA
701 ATACAGGTTA TCAATTTATT TTAAATCTCA TTGTTTCCAG TTAGCAATAT
751 CATACCTATT AAAGCTGTTC ATTGTAACAA AATTCAATCA AAAAGGCAGC
801 TAGGTGAGAA GGAACATAC CACTCTCATG GTTCATAGTA TTCACTGTAT
851 GTATGCTAGG GAAAAGACTT GCTCCAGTCT CCTCCTCAGT TCTGTGCCTG
901 AGAACCACTG CTGCATATAT TTGTTTTTAA ATTTTGTATT GAACTGTAA
951 TTGAAGCTTT AAAAGCATAT ATGAAATGTA TAAATCTAAG ATGTATAATA
1001 CATTATTGAC TCCAAAAAAA AAAAAA

```

BLAST Results

Entry HSG0101 from database EMBL:
human STS SHGC-35956.
Length = 401
Minus Strand HSPs:
Score = 1417 (212.6 bits), Expect = 9.3e-58, P = 9.3e-58
Identities = 301/311 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 40 bp to 573 bp; peptide length: 178
Category: strong similarity to known protein

```

1 MTTLRAFTCD DLFRFNNINL DPLTETYGIP FYLQYLAHP EYFIVAVAPG
51 GELMGYIMGK AEGSVAREEW HGHVTAHSV PEFRRGLAA KLMELLEIS

```

101 ERKGGFFVDL FVRVSNQVAV NMYKQLGYSV YRTVIEYYSA SNGEPDEDAY
151 DMRKALSRDT EKKSIIPLPH PVRPEDIE

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3g8, frame 1

TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4., N = 1, Score = 475, P = 3.2e-45

SWISSPROT:ARDH LEIDO N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 SUBUNIT HOMOLOG., N = 1, Score = 451, P = 1.1e-42

PIR:S69021 hypothetical protein YPR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 382, P = 2.3e-35

>TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4. Length = 180

HSPs:

Score = 475 (71.3 bits), Expect = 3.2e-45, P = 3.2e-45
Identities = 96/165 (58%), Positives = 118/165 (71%)

Query: 1 MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLAHWPEYFIVAVAPGGE--LMGYIM 58
MT R F DLF FNNINLDPLTET+ I FYL YL WP +V + + LMGYIM
Sbjct: 1 MTDTRKFKATDLFSFNNINLDPLTETFNISFYLSYLNKWPSPCVVQESDLSPTLMGYIM 60

Query: 59 GKAEGSVAREEWHGHVLTALSVAPEFRRLGLAAKLMELLEISERKGGFFVDLFVRVSNQV 118
GK+EG+ +EWH HVTAA++VAP RRLGLA +M+ LE + + FFVDLFVR SN +
Sbjct: 61 GKSEGT--GKEWHTHTAITVAPNSRRRLGLARTMDYLETVGNSENAFFVDLFVRASNAL 118

Query: 119 AVNMYKQLGYSVYRTVIEYYASNGEPDEDAYDMRKALSRDTEKKSII 165
A++ YK LGYSVYR VI YYS +G+ DED++DMRK LSRD ++SI
Sbjct: 119 AIDFYKGLGYSVYRRVIGYYSNPHGK-DEDSFDMRKPLSRDVRNRESI 164

Pedant information for DKFZphfbr2_3g8, frame 1

Report for DKFZphfbr2_3g8.1

[LENGTH] 178
[MW] 20338.24
[pI] 5.06
[HOMOL] TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4. 7e-47
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YPR131c] 6e-37
[FUNCAT] 01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] r general function prediction [M. jannaschii, MJ1530] 6e-09
[PIRKW] acyltransferase 1e-12
[SUPFAM] arrest-defective protein 1 1e-12
[SUPFAM] Escherichia coli peptide N-acetyltransferase rimI 1e-07
[PROSITE] CK2_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 3
[KW] Alpha_Beta

SEQ MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLAHWPEYFIVAVAPGGELMGYIMGK
PRD cccccccccchhhhhccccccccccccchhhhhccccceeeccccceeehhhh
SEQ AEGSVAREEWHGHVLTALSVAPEFRRLGLAAKLMELLEISERKGGFFVDLFVRVSNQVAV
PRD hccccccccceeehhhhhhhhcchhhhhhhhhhhhhccccceeeccccchhhhh
SEQ NMYKQLGYSVYRTVIEYYASNGEPDEDAYDMRKALSRDTEKKSIIPLPHPVRPEDIE
PRD hhhhhhccccchhhhhccccccccccccchhhhhhhhhhhhhhhhhhhcccccccccccc

Prosite for DKFZphfbr2_3g8.1

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	100->103	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	8->12	CK2_PHOSPHO_SITE	PDOC00006
PS00006	133->137	CK2_PHOSPHO_SITE	PDOC00006
PS00006	141->145	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphfbr2_3g8.1)

DKF2phfbr2_312

group: brain derived

DKF2phfbr2_312 encodes a novel 589 amino acid protein with weak similarity to *S. cerevisiae* ubiquitin-like protein DSK2.

Pfam predicts for this protein similarity to the ubiquitin family; No informative BLAST results; No predictive prosite or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ubiquitin-like protein DSK2 yeast

complete cDNA, complete cds, EST hits
Dsk2p is involved in spindle pole body SPB duplication, SPB = centomer
strong similarity to HRIHFB2157 human mRNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2978 bp

Poly A stretch at pos. 2958, polyadenylation signal at pos. 2924

```
1 GGGGGGAGGA AGCGGTGGCT GCTGCGGATG TCGGTGTGAG CGAGCGGGCG
51 CTGAACACAC GGGGGCTGCC GAGCGCCTGA CCCGGGCTTG CGCCAGAGCC
101 TGCACCGAGC TCCGGGGGCC CACACCCGCT ACGGTGGCCC TGCGCCCGTT
151 GCTACTGAGG CGGCGTGCTC TGCATTCTTC GCTGTCCAGG CCTGCCGGCT
201 CTGGTGTCTG CTGGCTCCTC CTTGCTCGCC TGCTCCCTCC TGCTTGCCCTG
251 AGTCACCGCC GCCGCCGCCG CCACAGCCAT GGCCGAGAGT GGTGAAAGCG
301 GCGGTCTCTC GGGCTCCCGG GATAGCGCCG CCGGAGCCGA AGGTGCTGGC
351 GCCCGCGCGG CCGCTGCCTC CGCGGAGCCC AAAATCATGA AAGTCACCGT
401 GAAGACCCCG AAGGAAAAGG AGGAATTCGC CGTGCCCGAG AATAGCTCCG
451 TCCAGCAGTT TAAGGAAGAA ATCTCTAAAC GTTTTAAATC ACATACTGAC
501 CAACTTGTGT TGATATTTGC TGGAAAAATT TTGAAAGATC AAGATACCTT
551 GAGTCAGCAT GGAATTCATG ATGGACTTAC TGTTACCTTT GTCATTAAAA
601 CACAAAACAG GCCTCAGGAT CATTACAGTC AGCAAACAAA TACAGCTGGA
651 GGCAATGTTA CTACATCATC AACTCCTAAT AGTAACTCTA CATCTGGTTC
701 TGCTACTAGC AACCCTTTTG GTTTAGGTGG CCTTGGGGGA CTTGCAGGTC
751 TGAGTAGCTT GGGTTTGAAT ACTACCAACT TCTCTGAATC ACAGAGTCAG
801 ATGCAGCGAC AACTTTTGTC TAACCCTGAA ATGATGGTCC AGATCATGGA
851 AAATCCCTTT GTTCAGAGCA TGCTCTCAA TCCGTACCTG ATGAGACAGT
901 TAATTATGGC CAATCCACAA ATGCAGCAGT TGATACAGAG AAATCCAGAA
951 ATTAGTCATA TGTGGAATAA TCCAGATATA ATGAGACAAA CGTTGGAATC
1001 TGCCAGGAAT CCAGCAATGA TGCAGGAGAT GATGAGGAAC CAGGACCGAG
1051 CTTTGAGCAA CCTAGAAAGC ATCCCAGGGG GATATAATGC TTTAAGGGCG
1101 ATGTACACAG ATATTCAGGA ACCAATGCTG AGTGCTGCAC AAGAGCAGTT
1151 TGGTGGTAAT CCATTGCTT CCTTGGTGAG CAATACATCC TCTGGTGAAG
1201 GTAGTCAACG TTCCCGTACA GAAATAGAG ATCCACTACC CAATCCATGG
1251 GCTCCACAGA CTTCCAGAGG TTCATCAGCT TCCAGCGGCA CTGCCAGCAG
1301 TGTGGGTGGC ACTACTGGTA GACTGCCCAG TGGCACTTCT GGGCAGAGTA
1351 CTACTGCGCC AAATTTGGTG CCTGGAGTAG GAGCTAGTAT GTTCAACACA
1401 CCAGGAATGC AGAGCTTGTT GCAACAAATA ACTGAAAACC CACAATGAT
1451 GCAAAACATG TTGTCTGCCC CCTACATGAG AAGCATGATG CAGTCACTAA
1501 GCCAGAATCC TGACCTTGCT GCACAGATGA TGCTGAATAA TCCCCTATTT
1551 GCTGGAATC CTCAGCTTCA AGAACAATG AGACAACAGC TCCCAACTTT
1601 CCTCCAACAA ATGCAGAAAT CTGATACACT ATCAGCAATG TCAAAACCTA
1651 GAGCAATGCA GGCCTTGTTA CAGATTCAGC AGGGTTTACA GACATTAGCA
1701 ACCGAAGCCC CGGGCCTCAT CCCAGGGTTT ACTCCTGGCT TGGGGGCATT
1751 AGGAAGCACT GGAGGCTCTT CGGGAACATA TGGATCTAAC GCCACACCTA
1801 GTGAAAACAC AAGTCCCACA GCAGGAACCA CTGAACCTGG ACATCAGCAG
1851 TTTATTGAGC AGATGCTGCA GGCTCTTGCT GGAGTAAATC CTCAGCTACA
1901 GAATCCAGAA GTCAGATTTC AGCAACAAT GGAACAATC AGTGCAATGG
1951 GATTTTGAAC CCGTGAAGCA AACTTGCAAG CTCTAATAGC AACAGGAGGT
2001 GATATCAATG CAGCTATTGA AAGTTACTG GGCTCCAGC CATCATAGCA
2051 GCATTTCTGT ATCTTGAAAA AATGTAATTT ATTTTGTATA ACGGCTCTTA
2101 AACTTTTAAA TACCTGCTTT ATTTCAATTT GACTCTTGGA ATTCTGTGCT
2151 GTTATAAACA AACCCTATAT GATGCATTTT AAGGTGGAGT ACAGTAAGAT
2201 GTTGGGTTT TTCTGTATTT TTCTTTCTG GAACAGTGGG AATTAAGGCT
2251 ACTGCATGCA TCACTTCTGC ATTTATTGTA ATTTTAAAA AACATCACCT
2301 TTTATAGTTG GGTGACCAGA TTTTGTCTG CATCTGTCCA GTTTATTTGC
2351 TTTTAAACA TTAGCCTATG GTAGTAATTT ATGTAGAATA AAAGCATTAA
2401 AAAGAAGCAA ATCATTTGCA CTCATAAAT TGTGGTACAG TATTGCTTAT
2451 TGTGACTTTG GCATGCATTT TTGCAACAA TGCTGTAAGA TTTATACTAC
2501 TGATAATTTT GTTTTATTTG TATACAATAT AGAGTATGCA CATTTGGGAG
```

```

2551 TGCATTCTCTG GAAACATACT GCAATAGGCT CTCTGAGCAA AACACCTGTA
2601 ACTAAAAAAG TGAAGATAAG AAAATACTCT TAAAGCTGAG TATTTCTTAA
2651 TTGTATAGAA TCTTACAGCA TCTTTGACAA ACATCTCCCA GCAAAAGTGC
2701 CGGTTAGTCA GGTTTGTTGA AAATACAGTA GAAAAGCTGA TTCTGGTTAT
2751 CTCTTTAAGG ACAATTAATT GTACAGACAC ATAATGTAAC ATTGTCTCAA
2801 CATTCAATCA CAGATTGACT GTAAATTACC TTAATCTTTG TGCAGACTGA
2851 AGGAACACTG TAGTATACCC CAAAGTGCAT TTGCCTAGGA CTTCTCAGCT
2901 TCTCCCATAG GTAGTTTAAC AGGCATTAAA ATTTGTAATT GAAATGTTGC
2951 TTCTACTCAA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 279 bp to 2045 bp; peptide length: 589
 Category: similarity to known protein

```

1 MAESGESGGP PGSQDSAAGA EGAGAPAAAA SAEPKIMKVT VKTPKEKEEF
51 AVPENSSVQQ FKEEISKRFK SHTDQLVLIF ACKILKDQDT LSQHGIDHGL
101 TVHLVIKTQN RPQDHSAAQT NTAGGNVTTT STPNNSNSTSG SATSNPFGLG
151 GLGGLAGLSS LGLNTTNFSE LQSOMQRQLL SNPEMMVQIM ENPFVQSMLS
201 NPDLMRQLIM ANPQMQLIQ RNPEISHMLN NPDIMRQTL LARNPAMMQE
251 MMRNQDRALS NLESIPGGYN ALRRMYTDIQ EPMLSAEQEQ FGGNPFASLV
301 SNTSSGEGSQ PSRTENRDPL PNPWAPQTSQ SSSASSGTAS TVGGTTGSTA
351 SGTSGQSTTA PNLVPGVGAS MFNTPGMQSL LQQITENPOL MQNMLSAPYM
401 RSMMSQLSQN PDLAAQMMLN NPLFAGNPQL QEOMRQQLPT FLQQMQNPDT
451 LSAMSNPRAM QALLQIQQGL QTLATEAPGL IPGFTPGLGA LGSTGSSSGT
501 NGSNATPSEN TSPTAGTTPE GHQQFIQQL QALAGVNPQL QNPEVRFQQQ
551 LEQLSAMGFL NREANLQALI ATGGDINAAI ERLLSQPS

```

BLASTP hits

Entry CE1_1 from database TREMBL:
 "F15C11.2"; Caenorhabditis elegans cosmid VF15C11L
 Length = 293
 Score = 454 (159.8 bits), Expect = 4.4e-43, P = 4.4e-43
 Identities = 81/162 (50%), Positives = 113/162 (69%)

Entry S54583 from database PIR:
 ubiquitin-like protein DSK2 - yeast (Saccharomyces cerevisiae)
 Length = 373
 Score = 278 (97.9 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 100/307 (32%), Positives = 155/307 (50%)

Entry AB015344_1 from database TREMBLNEW:
 gene: "HRIHFB2157"; Homo sapiens HRIHFB2157 mRNA, partial cds.
 Score = 1135, P = 3.6e-115, identities = 227/301, positives = 253/301

Alert BLASTP hits for DKFZphfbr2_312, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_312, frame 3

Report for DKFZphfbr2_312.3

```

[LENGTH]      589
[MW]           62489.22
[pI]           5.02
[HOMOL]        TREMBL:AB015344_1 gene: "HRIHFB2157"; Homo sapiens HRIHFB2157 mRNA, partial
cds. 1e-121
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YMR276w] 2e-17

```

```

[FUNCAT]      30.10 nuclear organization      [S. cerevisiae, YMR276w] 2e-17
[BLOCKS]      BL00299 Ubiquitin family proteins
[SUPFAM]      unassigned ubiquitin-related proteins 5e-16
[SUPFAM]      ubiquitin homology 5e-16
[PROSITE]     MYRISTYL 24
[PROSITE]     CK2_PHOSPHO_SITE 9
[PROSITE]     GLYCOSAMINOGLYCAN 1
[PROSITE]     PKC_PHOSPHO_SITE 3
[PROSITE]     ASN_GLYCOSYLATION 7
[PFAM]        Ubiquitin family
[KW]          Irregular
[KW]          3D
[KW]          LOW_COMPLEXITY 23.43 %

```

```

SEQ      MAESGESGGPPGSQDSAAGAEGAGAPAAAASAEPKIMKVTVKTPKEKEEFAVPENSSVQQ
SEG      ..xxxxxxxxxxxx..xxxxxxxxxxxxxxxxxxxx..xxxxxxxxxxxxx.....
laarA    .....CEEEEEETTTCEEEECTTTTBHHH

SEQ      FKKEISKRFKSHTDQLVLI FAGKILKDQDTLSQHGIDGLTVHLVIKTQNRPDHSAQQT
SEG      .....
laarA    HHHHHHHHHCCCGGEEEEETTEECTTTTBGGGGCCTTTTEEEEBE.....

SEQ      NTAGGNVTTSTPNSNSTSGSATSNPFGGLGGLAGLSSGLNNTNFSELQSQMQRQLL
SEG      ...xxxxxxxxxxxxxxxxxxxxxxxx..xxxxxxxxxxxxxxxxxxxx.....
laarA    .....

SEQ      SNPEMMVQIMENPFVQSMLSNPDLMRQLIMANPQMQLIQRNPEISHMLNPNPDIMRQLE
SEG      .....
laarA    .....

SEQ      LARNPAMMQEMMRNQDRALSNLIESIPGGYNALRRMYTDIQEPMLSAEQEFGGNPFASLV
SEG      .....
laarA    .....

SEQ      SNTSSGEGSQPSRTENRDPLPNPWAPQTSQSSSASSGTASTVGGTTGSTAGTSGQSTTA
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
laarA    .....

SEQ      PNLVPGVGASMFNTPGMQSLLOQITENPQLMQNMLSAPYMRSMMSLSQNPDLAAQMMLN
SEG      .....
laarA    .....

SEQ      NPLFAGNPQLQEQMRQLPTFLQMQNPDTLSAMSNPRAMQALLQIQGLQTLATEAPGL
SEG      .....
laarA    .....

SEQ      IPGFTPGLGALGSTGGSSGTNGSNATPSENTSPTAGTTEPGHQQFIQMLQALAGVNPQL
SEG      ....xxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
laarA    .....

SEQ      QNPEVRFQQLEQLSAMGFLNREANLQALITGGDINAATIERLLGSQPS
SEG      .....
laarA    .....

```

Prosites for DKF2phfbr2_312.3

PS00001	55->59	ASN_GLYCOSYLATION	PDOC00001
PS00001	126->130	ASN_GLYCOSYLATION	PDOC00001
PS00001	136->140	ASN_GLYCOSYLATION	PDOC00001
PS00001	164->168	ASN_GLYCOSYLATION	PDOC00001
PS00001	167->171	ASN_GLYCOSYLATION	PDOC00001
PS00001	302->306	ASN_GLYCOSYLATION	PDOC00001
PS00001	501->505	ASN_GLYCOSYLATION	PDOC00001
PS00002	305->309	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	40->43	PKC_PHOSPHO_SITE	PDOC00005
PS00005	43->46	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	71->75	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	200->204	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	572->576	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	12->18	MYRISTYL	PDOC00008

PS00008	19->25	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	140->146	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	162->168	MYRISTYL	PDOC00008
PS00008	267->273	MYRISTYL	PDOC00008
PS00008	293->299	MYRISTYL	PDOC00008
PS00008	308->314	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	492->498	MYRISTYL	PDOC00008
PS00008	495->501	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008

Pfam for DKF2phfbr2_312.3

HMM_NAME	Ubiquitin family		
HMM	*MQIFVKTLtGRTcTFEVepQEtVeqIKQHieekEGIPPeQQRLIFaGRQ		
	M ++VKT + +F V+++ V Q+K+ I+ +Q +LIFAG+		
Query	37	MKVTVKTPK-EKEEFVAVPENSSVQQFKEEISKRFKSHDQLVLIFAGKI	84
HMM	LEDeKTLsDYNiggeSTLHLVlR*		
	L D TLS+++I + T+HLV++		
Query	85	LKDQDTLSQHGIDGLTVHLVIK	107

DKFZphfbr2_62b11

group: signal transduction

DKFZphfbr2_62b11 encodes a novel 655 amino acid putative GTPase-activating protein, related to human chimaerins.

The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to CHIMAERIN

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="4"

Insert length: 4593 bp

Poly A stretch at pos. 4571, polyadenylation signal at pos. 4553

```

1  GGGGGAGTTT GAAGACAGAA AGGAAAGGGG AGAAACCTGC AGAGAGCATC
51 AAAGGATGGG GGGTGCTATA AAAGAAGCAG GGGGGTCCTT TGAAAGAAAT
101 CTATCATGCA CTGAAATGCT TTCTGGAGAA GGTGCCGTTA TTTTCCTCCC
151 CTCTTGCTCA GATGAAAGGA GCCAGCAAGG ACAGTCCTGA AATATTCCTC
201 AGGGGACTTT TTGTCAATGT TCCTCTTTCC TCTTGACAG AGCTATTTGC
251 TGACCTTTCC AGAGGAATCT CAGTCCAGCT GAGAAGACAG TTCTTAATAA
301 AAACAAAAAA ATGCAAAAAC CAATTCCTGC TGTTTGAATG GGAATGGTAG
351 CTTGCTTGCT GCAGTTCTTT TCCTGTGACA TTTTGGAATG TCTGCAGAAA
401 CTTAAAAAAA AGAAAAAAA AACCTTAAAA ACTCCCTGGA TTAGGCAAGA
451 GAAAGGAAGG TTTT TTTTCT TAAACAGGA GTAAATGAGA GGTGGTAACT
501 TATCCCTAAG CCAGGACCTG GATGATCAAA ACCTTCAAA TCTAGGGATC
551 AGCACTTCAA AATAACAAG TAAACAAGCA TGAGGAGTGG CTGTTGGGTT
601 TCGCTCAGAG GCAGGTTTAA AAGGAAGCCA AAACCGGGTT CAGAACTTCA
651 GGCCTGTACG ATGCCTGAAG ACCGGAATTC TGGGGGGTGC CCGGCTGGTG
701 CCTTAGCCTC AACTCCTTTC ATCCCTAAAA CTACATACAG AAGAATCAAA
751 CGGTGTTTAA GTTTTCGGAA AGGCATTTTT GGACAGAAAC TGGAGGATAC
801 TGTTTCGTAT GAGAAGAGAT ATGGGAACCG TCTGGCTCCG ATGTTGGTGG
851 AGCAGTGGCT GGACTTTATC CGACAAAGGG GGCTGAAAGA AGAGGGTCTC
901 TTTGCACTGC CAGGCCAGGC TAATCTTGT TAAAGAGCTCC AAGATGCCTT
951 TGACTGTGGG GAGAAGCCAT CATTGTACAG CAACACAGAT GTACACACGG
1001 TGGCATCACT TCTTAAGCTG TACCTCCGAG AACTTCCAGA ACCAGTTATT
1051 CCTTATGCGA AGTATGAAGA TTTT TTTTGTCA TGTGCCAAAC TGCTCAGCAA
1101 GGAAGAGGAA GCAGGTGTTA AGGAATTAGC AAAGCAGGTG AAGAGTTTGC
1151 CAGTGGTAAA TTACAACCTC CTCAGTATA TTTGCAGATT CTGGATGAA
1201 GTACAGTCTC ACTCGGGAGT TAACAAAATG AGTGTGCAGA ACTTGGCAAC
1251 GGTCTTTGGT CCTAATATCC TGCGCCCAAG AGTGGAGAT CTTTGACTA
1301 TCATGGAGGG CACTGTGGTG GTCCAGCAGT TGATGTCAGT GATGATTAGC
1351 AAACATGATT GCCTCTTTCC CAAAGATGCA GAACTACAAA GCAAGCCCA
1401 AGATGGAGTG AGCAACAACA ATGAAATTCA GAAGAAAGCC ACCATGGGGC
1451 TGTTACAGAA CAAGGAGAAC AATAACACCA AGGACAGCCC TAGTAGGCAG
1501 TGCTCCTGGG ACAAGTCTGA GTCACCCAG AGAAGCAGCA TGAACAATGG
1551 ATCCCCACA GCTCTATCAG GCAGCAAAAC CAACAGCCCA AAGAACAAGT
1601 TTCACAAGCT AGATGTGTCT AGAAGCCCCC CTCTCATGGT CAAAAGAAGC
1651 CCAGCCTTTA ATAAGGGTAG TGGGATAGTT ACCAATGGGT CCTTCAGCAG
1701 CAGTAATGCA GAAGGTCTTG AGAAAACCCA AACCACCCCC AATGGGAGCC
1751 TACAGGCCAG AAGGAGCTCT TCACTGAAGG TATCTGTGAC CAAAATGGGC
1801 ACGCACAGTG TACAGAATGG AACGGTGCGC ATGGGCATTT TGAACAGCGA
1851 CACACTCGGG AACCACCAA ATGTTGAAA CATGAGCTGG CTGCCAAATG
1901 GCTATGTGAC CCTGAGGGAT AACAAGCAGA AAGAACAAGC TGGAGAGTTA
1951 GGCCAGCACA ACAGACTGTC CACCTATGAT AATGTCCATC AACAGTTCTC
2001 CATGATGACA CTTGATGACA AGCAGAGCAT TGACAGTGCT ACCTGGTCCA
2051 CTTCTCTCTG TGAATCTCC CTCCCTGAGA ACTCCAATC CTGTCTGCTC
2101 TCTACCAACA CCTGCCAGA GCAAGACTTT TTTGGGGGGA ACTTTGAGGA
2151 CCCTGTTTTG GATGGGCCCC CGCAGGACGA CCTTTCCAC CCCAGGGACT
2201 ATGAAAGCAA AAGTGACCAC AGGAGTGTGG GAGGTCGAAG TAGTCGTGCC
2251 ACCAGTAGCA GTGACAACAG TGAGACATTT GTGGGCAACA GCAGCAGCAA
2301 CCACAGTGCA CTGCACAGTT TAGTTTCCAG CCTGAAACAG GAAATGACCA
2351 AACAGAAGAT AGAGTATGAG TCCAGGATAA AGAGCTTAGA ACAGCGAAAC
2401 TTGACTTTGG AAACAGAAAT GATGAGCCTC CATGATGAAC TGGATCAGGA
2451 GAGGAAAAAG TTCACAATGA TAGAAATAAA AATGCGAAAT GCCGAGCGAG
2501 CAAAGAGAAG TGCCGAGAAA AGAAATGACA TGCTACAGAA AGAAATGGAG
2551 CAGTTTTTTT CCACGTTTGG AGAACTGACA GTGGAACCCA GGAGAACCGA

```

```

2601 GAGAGGAAAC ACAATATGGA TTCAGTGAGC CTGCTTTCGC CTGCTGTCTC
2651 TGATGGCTCT GGCAAGGACT CCAGGGATTG TGGTGGGATA TGACTTAGAA
2701 CCAGGTGGCT GGTCACCTGG ATGTACAGAA GTCTAACTGG TGAAGGAATA
2751 TCATTACAG ACATTAAACA TCCATATCTG CAATGTGTAC CAAAGTTATA
2801 TCATGCCCCA TAATGCTACT GTCAAGTGT ACAACTGGAT ATGTGTATAT
2851 AGAGTAGTTT TTCAAAAGTA AACTAAAAAT GAGAAGCATA TTCAAGAAT
2901 TATTTTATG CAAGTCTTGT ATTTAAATGT TAAATCAATA TGTGTGTGCA
2951 ATTTAGCTTG CTTTCAAGCT TCACCCCTTG CACTTAACAT AAGCTATTTT
3001 TGGCATTGTG TTATCATCGG CTTATTTTAT AGATCAATAT TTTTATTTC
3051 CTTTTTGTCT GAGGAAATGA AGATAAGCAA AAATATAAAT ATATATATAA
3101 ATATATGAGT TATTTAAACC AGAAGAATAC TTTGTGGCTG TGCTGTTTGT
3151 GCCAATAGAC TTTGTCTATG CCAAAAAGAG AAATGTAAAT AGTTTATAA
3201 AATACAGTCG AATCACCAGG AACCTTTGAG CTGCTTTTAA AATTCTTCCC
3251 CTGGCACCAC TCAGTTTTCG TTTTGCGAGG CGATTTGACA TAGGAACTTT
3301 GAGACTCCAT GAGAAAGTCC CTTTCTGAGG CCCACTGTCT ACCTTGCCAG
3351 ATCCTCAGTG CGTATCGCCA ATGCAGGATG CTCCTTAGAA AAGAAAAAAT
3401 GGTAAAGGAT GGCATTTAAC GATTCAAGCT TTGAATTACT CTGCTCCCTCT
3451 GGACCGAATC TCTTTAACTG CTGGATAGTT TTAGAGGAAT TCTCCTGCTA
3501 CTTAGGTACT GGGAAACAAT GCTTGCTAAA CCATGCCAC GTGAGCACCT
3551 GTCTCCCACT CAAACCTCTC CCATCTCCCA ACAACTGCAC TTAGAATAC
3601 CAGCAGTGAA ATGGTATTAC TGTTTCCCTC TGAGTGAAAC TGCTAGAGTA
3651 TATGTCACGT AGTGACATTT TTTTCTCACT CAGGCTATTG CCATCTGGGA
3701 TTCTCTCCCT ACTACAGCTG GCAAAGTTGG TTTGCAGCAA GAAGATAGTG
3751 GGAGGGGGCC AGGCTGCAGG AGAAGGAGAA AAGTTTAGAA GAAACAAACC
3801 ATTTTGTCTT TAATTTTGAC AGTATCACTT TCCTGTATAA ACATACAATA
3851 ATTTTAAAG GTGAATGCCT AAAGTTCCAA TTTTAGCAA TATGGGAACC
3901 TCAGCAATGC TAATTTTCTA GAAAACCCA GGGCTCTTTG GAGCTAGAGT
3951 TTTGGGAGAA CAGTTCTTCA CAATAAGGCA ATGGTTTGA GAGGCCAGGC
4001 AAATAATCTT TCTCACCCTA GAACAAAAG TTACAAAAGG CATAATCGGA
4051 AATAGAGACT ACATACTTGA GTTATGGGG TTTGTGTGT TGAAGGTTT
4101 AATGCTTGCA TGTGTTTATT TATTTTCAAG AGGGAAAGTG GTCTGTACTG
4151 CTTTCATCCT TGCCACTGTC TTGCTTTTAT TTTTACTCT CCCACTGAGC
4201 AAGCGTCTGT GGTCCATGG TATCAACCAG TATCTTTATA GCAATAATTT
4251 CTTTAATTC CTTTCTCTC TCTTCCAAT TATTTAACC GTTACTTCCA
4301 CCTGGACATA CGATAGGAAA TTCAACTCA AAATATGAAA ATTGATCTTA
4351 ATAACCTCTC CTTCATATCT TTTCACCTAT TTCCAGTCTT TATCATAGTT
4401 GATAAAACC TCAGACTCAT CCAGAAAGCT ATATGATGCA CTAGTAAAAA
4451 AAACAAAGAT ATTTAAACTG CTTGGGTTCA AATGGTATAC AATTTGCCAG
4501 CTGTTACTGA ACCTTCTATG CATAACTTTT TTTTCTCTCT GTGCAATTGG
4551 AATAATAAAA ATACTACTCC CATAAAAAA AAAAAAAA AAC

```

BLAST Results

Entry G38474 from database EMBLNEW:
 SHGC-58303 Human Homo sapiens STS genomic, sequence tagged site.
 Score = 2175, P = 1.2e-92, identities = 439/441

Medline entries

97476250:
 Beta2-chimaerin is a high affinity receptor for the phorbol ester tumor promoters.

Peptide information for frame 1

ORF from 661 bp to 2625 bp; peptide length: 655
 Category: similarity to known protein

```

1 MPEDRNSGGC PAGALASTPF IPKTTYRRIK RCFSFRKGIF GQKLEDTVRY
51 EKRYNRLAP MLVEQCVDPI RQRLKEEGL FRLPGQANLV KELQDAFDCG
101 EKPSFDSNTD VHTVASLLKL YLRELPEPVI PYAKGYEDFLS CAKLLSKEEE
151 AGVKELAKQV KSLPVVNNYL LKYICRFLDE VQSYSGVNKM SVQNLATVFG
201 PNILRPKVED PLTIMEGTVV VQQLMSVMIS KHDCLFPKDA ELQSKPQDGV
251 SNNNEIQKKA TMGLLQNKEN NNTKDSPSRQ CSWDKSESPQ RSSMNGSPT
301 ALSGSKTNSP KNSVHKLDVS RSPPLMVKKN PAFNKGSGIV TNGSFSSSNA
351 EGLEKTQTFP NGSLOARRSS SLKVSQTKMG THSVQNGTVR MGILNSDTLG
401 NPTNVRNMSW LPNGYVTLRD NKQKEQAGEL GQHNRLSTYD NVHQQFSMMN
451 LDDKQSIDSA TWSTSSCEIS LPENSNSCRS STTTCPEQDF FGGNFEDPVL
501 DGPPQDDLH PRDYESKSDH RSVGGRSSRA TSSSDNSETF VGNSSSNHSA
551 LHSLVSSLKQ EMTKQKIEYE SRIKSLEQRN LTLETMMMSL HDELDQERKK

```

601 FTMIEIKMRN AERAKEDA EK RNDMLQKEME QFFSTFGELT VEPRRTERGN
651 TIWIQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62b11, frame 1

SWISSPROT:Y053 HUMAN HYPOTHETICAL PROTEIN KIAA0053., N = 3, Score = 661, P = 2.4e-89

TREMBL:HSU90908_1 product: "unknown"; Human clones 23549 and 23762 mRNA, complete cds., N = 1, Score = 348, P = 1.1e-29

PIR:S29128 N-chimerin - rat, N = 1, Score = 286, P = 2.8e-24

PIR:S29956 beta-chimerin - rat, N = 1, Score = 279, P = 1.6e-23

TREMBL:AB014572_1 gene: "KIAA0672"; product: "KIAA0672 protein"; Homo sapiens mRNA for KIAA0672 protein, complete cds., N = 1, Score = 314, P = 1e-24

>SWISSPROT:Y053 HUMAN HYPOTHETICAL PROTEIN KIAA0053.
Length = 638

HSPs:

Score = 661 (99.2 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 122/209 (58%), Positives = 160/209 (76%)

Query: 38 GIFGQKLEDTVRYEKRYGNRLAPMLVEQCVD FIRQRGLKEEGLFRLPGQANLVKELQDAF 97
G+FGQ+L++TV YE+++G L P+LVE+C +FI + G EEG+FRLPGQ NLVK+L+DAF
Sbjct: 148 GVFGQRLDETVAVEQKFGPHLVPIVLEKCAEFIEHGRNEEGIFRLPGQDNLVKQLRDAF 207

Query: 98 DCGEKPSFDSNTDVHTVASLLKLYLRELPEPVIPIYAKYEDFLSCAKLLSKEEEAGVKELA 157
D GE+PSFD +TDVHTVASLLKLYLR+LPEPV+P+++YE FL C +L + +E +EL
Sbjct: 208 DAGERPSFDRD TDVHTVASLLKLYLRDLPEPVVPSQYEGFLLCGQLTNADEAKAQQELM 267

Query: 158 KQVKSLPVVNYNLLKYICRFLDEVQSYSGVNMKSQNLATVFGPNILRPKVEDPLTIMEG 217
KQ+ LP NY+LL YICRFL E+Q VNMKSV NLATV G N++R KVEDP IM G
Sbjct: 268 KQLSILPRDNYSLSYICRFLHEIQLNCAVNMKSVDNLATVIGVNLIRSKVEDPAVIMRG 327

Query: 218 TVVVQQLMSVMISKHDCFLFPKDAELQSKP 246
T +Q++M++MI H+ LFPK ++ P
Sbjct: 328 TPQIQRVMTMMIRDHEVLFPKSKDIP LSP 356

Score = 210 (31.5 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 45/115 (39%), Positives = 73/115 (63%)

Query: 531 TSSSDNSETFVGNSSSNHSLHSL---VSSLKQEMTKQKIEYESRIKSLEQRNL TLET 587
T +S NSET G +S + SL V L++E+ QK YE +IK+LE+ N + ++
Sbjct: 523 TLASPNSETGPGKKNSGEEIDSLQRMVQELRKEIETQKQMYEEQIKNLEKENYDVWAKV 582

Query: 588 MSLHDELDQERKKFTMIEIKMRNAERAKEDA EK RNDMLQKEME QFFSTFGELTVE 642
+ L++EL++E+KK +EI +RN ER++ED EKR N L++E++F + E E
Sbjct: 583 VRLNEELEKEKKKSAALEISLRNMERSREDVEKRNKALEEEVKEFVKSMKEPKTE 637

Score = 70 (10.5 bits), Expect = 1.2e-74, Sum P(3) = 1.2e-74
Identities = 28/121 (23%), Positives = 54/121 (44%)

Query: 528 SRATSSSDNSETFVGNSSSNHSLHSLVSSLKQE-MTKQKIEYESRIKSLEQRNL-TLET 585
S+ TS+ DN + G+ SAL S K + + E K+ + + +L+
Sbjct: 489 SQRTSTYDNVPSLPGSPGEEASALSSQACDSKGD TLASPNSETGPGKKNSGEEIDSLQR 548

Query: 586 EMMSLHDELDQERKKFTMIEIKMRNAERAKEDA EK RNDMLQKEME QFFSTFGELTVEPRR 645
+ L E++ +++ ME ++N E+ D + L +E+E+ L + R
Sbjct: 549 MVQELRKEIETQKQ---MYEEQIKNLEKENYDVWAKVVRLNEELEKEKKKSAALEISLRN 605

Query: 646 TER 648
ER
Sbjct: 606 MER 608

Score = 53 (8.0 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 31/111 (27%), Positives = 46/111 (41%)

Query: 344 SFSSSNAEGLEKTQTPNGSLQARRSSSLKVS GMTKMGTHSVQNG----TV--RMGILNSD 397
SFSS ++ + T T A S KV K G +Q+ T+ R L S
Sbjct: 388 SFSSMTSDS-DTTSPTGQQPSDAFPEDSSKVPREKPGDWKMQSRKRTQTLPNRKCF LSA 446

Query: 398 TLG-NPTNV---RNMSWLPNGYVTLRDNKQKEQAGELGQ---HNRLSTYDNV 442
 G N + + +N W P+ + ++ + +L Q R STYDNV
 Sbjct: 447 FQGANSKMEIFKNEFWSPSSEAKAGEGHRRTMSQDLRQLSDSQRTSTYDNV 498

Score = 53 (8.0 bits), Expect = 3.5e-14, Sum P(3) = 3.5e-14
 Identities = 32/125 (25%), Positives = 56/125 (44%)

Query: 242 LQSKPQDG---VSNNEIQKKATMGLLNQKEN--NNTKD---SPSRQCSWDKSESQQRSS 293
 ++SK +D + +IQ+ TM ++++ E +KD SP Q + K RSS
 Sbjct: 314 IRSKVEDPAVIMRGTPQIQRVMTM-MIRDHEVLFPKSKDIPLSPPAQKNDPKKAPVARSS 372

Query: 294 MNNGSPTALSGSKTNSPKNSVHKLDVSRSPPLMVKKNPAPFNKSGSIVTNGSFSSSSNAEGL 353
 + + L S+T+S + D + P + + AF + S V +
 Sbjct: 373 VGWDATEDLRISRTDSFSSMTSDSDTTS--PTGQQPSDAFPEDSSKVPREKPGDWKMQSR 430

Query: 354 EKTQTPN 361
 ++TQT PN
 Sbjct: 431 KRTQTLPN 438

Pedant information for DKFZphfbr2_62b11, frame 1

Report for DKFZphfbr2_62b11.1

[LENGTH] 655
 [MW] 73394.60
 [pI] 8.13
 [HOMOL] SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053. 3e-71
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YPL115c] 1e-16
 [FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 1e-16
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YPL115c]
 1e-16
 [FUNCAT] 10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 1e-16
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-16
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-16
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YDR379w] 4e-16
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-15
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 2e-13
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 2e-13
 [SCOP] dlrp_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens)] 2e-46
 [SCOP] dlpbwa_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Homo sapiens)] 6e-37
 [PIRKW] phosphotransferase 3e-13
 [PIRKW] breakpoint cluster region 2e-20
 [PIRKW] transmembrane protein 7e-14
 [PIRKW] brain 2e-20
 [PIRKW] alternative splicing 2e-20
 [PIRKW] P-loop 9e-19
 [PIRKW] cytoskeleton 1e-08
 [SUPFAM] CDC24 homology 7e-21
 [SUPFAM] bcr protein 7e-21
 [SUPFAM] myosin motor domain homology 9e-19
 [SUPFAM] pleckstrin repeat homology 2e-15
 [SUPFAM] LIM metal-binding repeat homology 9e-15
 [SUPFAM] protein kinase C zinc-binding repeat homology 5e-24
 [PROSITE] MYRISTYL 16
 [PROSITE] CAMP_PHOSPHO_SITE 3
 [PROSITE] CK2_PHOSPHO_SITE 15
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 11
 [PROSITE] ASN_GLYCOSYLATION 8
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 6.87 %
 [KW] COILED_COIL 12.06 %

SEQ MPEDRNSGGCPAGALASTPFIPKTTYRRIKRCFSFRKGIFGQKLEDTVRYEKRYGNRLAP
 SEG
 COILS
 lrgp-C

SEQ MLVEQCVDVFIQRGLKEEGLFRLPGQANLVKELQDAFDCGEKPSFDSNTDVHTVASLLKL
 SEG
 COILS
 lrgp- HHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHCCCCGGGCCCHHHHHHHHHH

SEQ YLRELPEPVIPIYAKYEDFLSCAKLLSKEEEAGVKELAKQVKSIPVVNYNLLKYICRFLDE
 SEG

```

COILS .....
1rgp- HHHHTTTTTTGGGHHHHHH--TTTTCGGGHHHHHHHHHHCCHHHHHHHHHHHHHHHH

SEQ      VQSYSGVNKMSVQNLATVFGPNILRPKVEDPLTIMEGTVVVQQLMSVMISKHDCLFPKDA
SEG      .....
COILS    .....
1rgp-    HHHHHHHHCCCHHHHHHHHGGGCC.....

SEQ      ELQSKPQDGVSNNEIQKKATMGLLQNKENNNTKDSPSRQCSWDKSESQRRSSMNGSPT
SEG      .....
COILS    .....
1rgp-    .....

SEQ      ALSGSKTNSPKNSVHKLDVSRSPPLMVKKNPAPFNKGSGIVTNGSFSSSNAEGLEKTQTP
SEG      .....
COILS    .....
1rgp-    .....

SEQ      NGSLQARRSSSLKVSCTKMGTHSVQNGTVRMGILNSDTLGNPTNVRNMSWLPNGYVTLRD
SEG      .....
COILS    .....
1rgp-    .....

SEQ      NKQKEQAGELGQHNRLSTYDNVHQQFSMMNLDDKQSIDSATWSTSSCEISLPENSNSCRS
SEG      .....
COILS    .....
1rgp-    .....

SEQ      STTTCPEQDFGNGFEDPVLGPPQDDLSPRDYESKSDHRSVGGRRSSRATSSSDNSETF
SEG      .....
COILS    .....
1rgp-    .....

SEQ      VGNSSSNHSLVSSSLKQEMTKQKIEYESRIKSLEQRNLTLETMMSLHDELDQERKK
SEG      .....
COILS    .....
1rgp-    .....

SEQ      FTMIEIKMNAERAKEDAERNDMLQKEMEQQFFSTFGELTVEPRRTERGNTIWIQ
SEG      .....
COILS    .....
1rgp-    .....

```

Prosites for DKFZphfbr2_62b11.1

PS00001	271->275	ASN_GLYCOSYLATION	PDOC00001
PS00001	342->346	ASN_GLYCOSYLATION	PDOC00001
PS00001	361->365	ASN_GLYCOSYLATION	PDOC00001
PS00001	386->390	ASN_GLYCOSYLATION	PDOC00001
PS00001	407->411	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	547->551	ASN_GLYCOSYLATION	PDOC00001
PS00001	580->584	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	367->371	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	599->603	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	47->50	PKC_PHOSPHO_SITE	PDOC00005
PS00005	309->312	PKC_PHOSPHO_SITE	PDOC00005
PS00005	371->374	PKC_PHOSPHO_SITE	PDOC00005
PS00005	388->391	PKC_PHOSPHO_SITE	PDOC00005
PS00005	417->420	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	527->530	PKC_PHOSPHO_SITE	PDOC00005
PS00005	557->560	PKC_PHOSPHO_SITE	PDOC00005
PS00005	646->649	PKC_PHOSPHO_SITE	PDOC00005
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	213->217	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	348->352	CK2_PHOSPHO_SITE	PDOC00006
PS00006	417->421	CK2_PHOSPHO_SITE	PDOC00006
PS00006	437->441	CK2_PHOSPHO_SITE	PDOC00006
PS00006	465->469	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	516->520	CK2_PHOSPHO_SITE	PDOC00006
PS00006	532->536	CK2_PHOSPHO_SITE	PDOC00006

PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	602->606	CK2_PHOSPHO_SITE	PDOC00006
PS00006	635->639	CK2_PHOSPHO_SITE	PDOC00006
PS00007	43->51	TYR_PHOSPHO_SITE	PDOC00007
PS00007	176->185	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	263->269	MYRISTYL	PDOC00008
PS00008	297->303	MYRISTYL	PDOC00008
PS00008	304->310	MYRISTYL	PDOC00008
PS00008	338->344	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	376->382	MYRISTYL	PDOC00008
PS00008	392->398	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	542->548	MYRISTYL	PDOC00008

{No Pfam data available for DKFZphfbr2_62b11.1}

DKF2phfbr2_62f10

group: intracellular transport and trafficking

DKF2phfbr2_62f10 encodes a novel 320 amino acid protein with strong similarity to mammalian zinc transporter proteins.

The novel proteins is a membrane protein, which should be involved in the transport of Zinc across the cell membrane.

The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide.

The new protein can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation.

strong similarity to zinc transporter proteins ;

membrane regions: 5

Summary DKF2phfbr2_62f10 encodes a novel 320 amino acid protein with similarity to zinc transporter protein.

The new protein can find clinical application in modulating Zn²⁺ uptake.

strong similarity to zinc transporter proteins

complete cDNA, complete cds, few EST hits

Sequenced by LMU

Locus: unknown

Insert length: 5422 bp

Poly A stretch at pos. 5397, polyadenylation signal at pos. 5381

```

1  GTCTAACTTT  GGAAATATCA  CCCTCATGCT  GTCTTCCCAG  GATGTCTCTC
51  TCCCTAAGTA  AGGGATGTTA  CTTCCTGGAG  GGAATGCAGT  GTTGGGAATC
101  TGAAGACCCA  GCTTTGAGCT  GAATTTGCTT  TGTGATACCT  GGAGAGAAGA
151  CGTGTTTTCT  TGACAACAGC  ACAGTACCTA  GTGAGTTCAA  CAACAACGAC
201  AACACAGCC  GCAGCTCATC  CTGGCCGTCA  TGGAGTTTCT  TGAAAGAGCG
251  TATCTTGTGA  ATGATAAAGC  TGCCAAGATG  TATGCTTTCA  CACTAGAAAG
301  AAGGAGCTGC  AAATGAACAC  TTCATAGCAA  TGTGGAATC  CAACAGAAAC
351  CGGTGAATAA  AGATCAGTGT  CCCAGAGAGA  GACCAGAGGA  GCTGGAGTCA
401  GGAGGCATGT  ACCACTGCCA  CAGTGGCTCC  AAGCCACAG  AAAAGGGGGC
451  GAATGAGTAC  GCCTATGCCA  AGTGGAAACT  CTGTTCTGCT  TCAGCAATAT
501  GCTTCATTTT  CATGATTGCA  GAGGTCGTGG  GTGGGCACAT  TGCTGGGAGT
551  CTTGCTGTGT  TCACAGATGC  TGCCACCTC  TTAATTGACC  TGACCAGTTT
601  CCTGCTCAGT  CTCTTCTCCC  TGTGGTTGTC  ATCGAAGCCT  CCCTCTAAGC
651  GGCTGCATT  TGGATGGCAC  CGAGCAGAGA  TCCTTGGTGC  CCTGCTCTCC
701  ATCCTGTGCA  TCTGGGTGGT  GACTGGCGTG  CTAGTGATAC  TGGCATGTGA
751  GCGCCTGCTG  TATCCTGATT  ACCAGATCCA  GCGCACTGTG  ATGATCATCG
801  TTTCCAGCTG  CGCAGTGCG  GCCAACATTG  TACTAACTGT  GGTTTTGCAC
851  CAGAGATGCC  TTGGCCACAA  TCACAAGGAA  GTACAAGCCA  ATGCCAGCGT
901  CAGAGCTGCT  TTTGTGCATG  CCCCTGGAGA  TCTATTTTCA  AGTATCAGTG
951  TGCTAATTAG  TGCATTATT  ATCTACTTTA  AGCCAGAGTA  TAAAATAGCC
1001  GACCCAATCT  GCACATTCAT  CTTTCCATC  CTGGTCTTGG  CCAGCACCAT
1051  CACTATCTTA  AAGGACTTCT  CCATCTTACT  CATGGAAGGT  GTGCCAAGA
1101  GCCTGAATTA  CAGTGGTGTG  AAAGAGCTTA  TTTTAGCAGT  CGACGGGGTG
1151  CTGTCTGTGC  ACTGCCTGCA  CATCTGGTCT  CTAACAATGA  ATCAAGTAAT
1201  TCTCTCAGCT  CATGTTGCTA  CAGCAGCCAG  CCGGGACAGC  CAAGTGGTTC
1251  GGAGAGAAAT  TGCTAAAGCC  CTTAGCAAAA  GCTTTACGAT  GCACTCACTC
1301  ACCATTACCA  TGGAAATCTC  AGTTGACCAG  GACCCCGACT  GCCTTTTCTG
1351  TGAAGACCCC  TGTGACTAGC  TCAGTCACAC  CGTCAGTTTC  CCAAAATTGA
1401  CAGGCCACCT  TCAAACATGC  TGCTATGCAA  TTTCTGCATC  ATAGAAAATA
1451  AGGAACCAAA  GGAAGAAAT  CATGTCATGG  TGCAATGCAT  ATTTTATCTA
1501  TTTATTTAGT  TCCATTCACC  ATGAAGGAAG  AGGCACTGAG  ATCCATCAAT
1551  CAATTGGATT  ATATACTGAT  CAGTAGCTGT  GTTCAATTGC  AGGAATGTGT
1601  ATATAGATTA  TTCCTGAGTG  GAGCCGAAGT  AACAGCTGTT  TGTAACATATC
1651  GGCAATACCA  AATTCTATCT  CCTTCCAATA  ATGCATCTTG  AGAACACATA
1701  GGTAATTTG  AACTCAGGAA  AGTCTTACTA  GAAATCAGTG  GAAGGGACAA
1751  ATAGTCACAA  AATTTTACCA  AAACATTAGA  AACAAAAAAT  AAGGAGAGCC
1801  AAGTCAGGAA  TAAAAGTGAC  TCTGTATGCT  AACGCCACAT  TAGAAGTTGG

```

```

1851 TTCTCTCACC AAGCTGTAAT GTGATTTTTT TTTCTACTCT GAATTGGAAA
1901 TATGTATGAA TATACAGAGA AGTGCTTACA ACTAATTTTT ATTTACTTGT
1951 CACATTTTGG CAATAAATCC CTCTTATTTT TAAATTTCTAA CTTGTTTATT
2001 TCAAAACTTT ATATAATCAC TGTTCAAAAG GAAATATTTT CACCTACCAG
2051 AGTGCTTAAA CACTGGCACC AGCCAAAGAA TGTGGTTGTA GAGACCCAGA
2101 AGTCTTCAAG AACAGCCGAC AAAACATTC GAGTTGACCC CACCAAGTTG
2151 TTGCCACAGA TAATTTAGAT ATTTACCTGC AAGAAGGAAT AAAGCAGATG
2201 CAACCAATTC ATTCAGTCCA CGAGCATGAT GTGAGCACTG CTTTGTGCTA
2251 GACATTTGGG TTAGCACTGA AACTATAAAG AGGAATCAGA CGCAGCAAGT
2301 GCTTCTGTGT TCTGGTAGCA ACTCAACACT ATCTGTGGAG AGTAAACTGA
2351 AGATGTGCAG GCCAACATTC TGGAAATCCT ATGTCAGTGG GTTTGGTTTG
2401 GAACCTGGAC TTCTGCATTT TTAAGAGTTA CCCAGAGATG CTTCTAAAGA
2451 TGAGCCATAG CTAGAAGAT TGTCAACCAC AGGAGTTTAT TGAGTGGGAC
2501 AGCTAGACAC ATACATTGGC AGTTACAATA GTATCATGAA TTGCAATGAT
2551 GTAGTGGGGT ATAAAGGAA AGCGATGGAT ATTGCCGGAT GGGCATGGCC
2601 AGTGATGTTT CACGTCATTG AGGTGACAGC TCTGCTGGAC TTTGAATTAC
2651 ATATGGAGGC TCTCCAGGAA GACGAAGAAG AGAAGGACAT TCTAGGCAAA
2701 AAGAAGACTA GGCACAAGGC ACACCTATGT TTGTCTGTTA GCTTTTAGTT
2751 GAAAAAGCAA AATACATGAT GCAAGAAAC CTCTCCACGC TGTGATTTTT
2801 AAAACTACAT ACTTTTGCA ACTTTATGGT TATGAGTATT GTAGAGAACA
2851 GGAGATAGGT CTTAGATGAT TTTTATGTTG TTGTCAGACT CTAGCAAGGT
2901 ACTAGAAACC TAGCAGGCAT TAATAATTGT TGAGGCAATG ACTCTGAGGC
2951 TATATCTGGG CCTTGTCAAT ATTTATCATT TATATTTGTA TTTTTCCTG
3001 AAATTTGAGG GCCAAGAAA CATTGACTTT GACTGAGGAG GTCACATCTG
3051 TGCCATCTCT GCAAAATCAAT CAGCACCAC T GAAATAACTA CTTAGCATT
3101 TGCTGAGGTT TCCCTGCTCA GTAGAGACAA ATATACTCAT CCCCCACCTC
3151 AGTGAGCTTG TTTAGGCAAC CAGGATTAGA GCTGCTCAGG TTCCCAACGT
3201 CTCCTGCCAC ATCGGGTTCT CAAAATGGAA AGAATGGTTT ATGCCAAATC
3251 ACTTTTCTCG TCTGAAGGAC CACTGAATGG TTTTGTGTTT CCATATTTTG
3301 CATAGGACGC CCTAAAGACT AGGTGACTTG GCAACACAC AAGTGTAGT
3351 ATAATCTTTT GCTTCTGCTT CTTTTGAAA ATCATGTTTA GATTTGATTT
3401 TAAGTCAGAA ATTCATGAA TGTCAGGTAA TCATTATGGA GGGAGATTTG
3451 TGTTGCAACC AAAGTAATTG TCCCATGGCC CCAGGGTATT TCTGTTGTTT
3501 CCCTGAAATT CTGCTTTTTT AGTCAGCTAG ATTGAAAAC CTGAACAGTA
3551 GATGTTTATA TGGCAAAATG CAAGACAATC TATAAGGGAG ATTTTAAGGA
3601 TTTTGAGATG AAAAAACAGA TGCTACTCAG GGGCTTTATG GACCATCCAT
3651 CAATTCTGAA GTTCTGATC TCCCATACC CTTTCCCTGG TGTGGTCAGA
3701 ACTCCAGTGC ACTGGAAGTT AGTGAATCA TGTAGTTGAA TTCTTTACTT
3751 CAAGACATTG TATTCTCTCC AGCTATCAAA ACATTAATGA TCTTTTATGT
3801 CTTTTTTTTT TTATTGTTAT ACTTTAAGTT CTGGGGTACA TGTGCGGAAC
3851 ATGTAGGTTT GTTACATAGG TATACATGTG CCATGGTGGT TTGCTGCACT
3901 CATCAACCTG TCATCTACAT TCTTTTATGT CTGCTTTTCA AAGCAACACT
3951 CTGTTCTTCT GAGTAGTGAA ATCAGGTCAA CTTTACCACC AGCCTCCATT
4001 TTTAATATGC TTCACCATCA TCCAGCACCT ACTTAAGATT TATCTAGGGC
4051 TCTGTGGTGA TGTTAGGACC CATAAAGAA ATTTATGCCT TCCATATGTT
4101 TGGTTACAGA TGGGAAATGG GAATGTTGAA GGACATGAAA GAAAGGATGT
4151 TTACACATTA AGCATCAGTT CTGAAGCTAG ATTGTCTGAG TTTGAATCTT
4201 AGCTCTTCCC TTTATTAGCT CTGTGACCTC GAGCTAGTTA CTTAAATGCT
4251 CTGATCCTCT ATTTCCCTGAT CAGTGAACC TCCCTATTCA AATGTGTGAG
4301 AGTTTAATAA ATTAGGACAC TTAATAATGT TGGAGCAGTG CATAGCATGT
4351 AGTGTTCACT ACATGTTAAA TGTGTTTTT TATTATGTAC AAACATGTGT
4401 GGGCAGAGAA TTTTAAATCA TCTCAACTTT TGAGAAATTT TGAGTTATCA
4451 ACACCGTTCC CACAAGACAG TGGCAAAATT ATTGGTGAGA ATTAACAGC
4501 TGTTTCTCAG AGGAAGCAAT GGAGGCTTGC TGGGATAAAG GCATTTACTG
4551 AGAGGCTGTT ACCTAGTGAG AGTGATGAAT TAATTAAAT AGTCGAATCC
4601 CTTTCTGACT GTCTCTGAAA GCTTCCGCTT TTATCTTTGA AGAGCAGAAT
4651 TGTACCCCCA AGGACATTTA TTAATAAAAA GAACAACGT CCAGTGCAAT
4701 GAAGGCAAG TCATAGGTCT CCCAAGTCTT ACCCCATTCC TGTGAAATAT
4751 CAAGTTCTTG CTTTCTCTCT GTCATGTAGC CTCAACTTTC TCCGACCGGG
4801 TGCATTTCTT TCTCTGGTTT CTAATTTGCC AGTGGCAAAT TTGGATCACT
4851 TACTTAATAT CTGTTAAATT TTGTGACCCA ACAAGTCTT TTAGCACTGT
4901 GGTGTCAAAA AGAAAAACAC CTCACAGGCA TATACATTT ATAGATTCCT
4951 GGAGAATGTT GCTCTCCAGC TCCATCCCCA CCCAATGAAA TATGATCCAG
5001 AGAGTCTTGC AAAAGAGACAA GCCTCATTTT CCACAATTAG CTCTAAAGTG
5051 CCTCCAGGAA ATGATTTTCT CAGCTCATCT CTCTGTATTC CCTGTTTTGG
5101 ATCACAGGCG AATCTGTTTA AATGACTAAT TACAGAAATC ATTAAGGCA
5151 CCAAGCAAAAT GTCATCTCTG AATACACACA TCCCAAGCTT TACAAATCCT
5201 GCCTGGCTTG ACAGTGATGA GGCCACTTAA CAGTCCAGCG CAGGCGGATG
5251 TTAATAAAAA TAAAAAGGTG ACCATCTGCG GTTTAGTTT TTAACTTTCT
5301 GATTTACAC TTAACGTCTG TCATTCTGTT ACTGGGCACC TGTTTAAATT
5351 CTATTTTAAA ATGTTAATGA GTGTTGTTTA AAATAAAATC AGGAAAGAGA
5401 GAAAAA AAAA AC

```

BLAST Results

No BLAST result

Medline entries

97121493:
ZnT-3, a putative transporter of zinc into synaptic vesicles.

96203098:
ZnT-2, a mammalian protein that confers resistance to zinc by
facilitating vesicular
sequestration.

Peptide information for frame 2.

ORF from 407 bp to 1366 bp; peptide length: 320
Category: strong similarity to known protein

```

1 MYCHSGSKP TEKGANEYAY AKWKLCSASA ICFIFMIAEV VGGHIAGSLA
51 VVTDAAHLLI DLTSFLLSLF SLWLSSKPPS KRLTFGWHRA EILGALLSIL
101 CIWVVTGVLV YLACERLLYP DYQIQATVMI IVSSCAVAAN IVLTVVLHQR
151 CLGHNHKEVQ ANASVRAAFV HAPGDLFQSI SVLISALIIY FKPEYKIADP
201 ICTFIFSIIV LASTITILKD FSILLMEGVP KSLNYSVGVE LILAVDGVLS
251 VHCLHIWSLT MNQVILSAHV ATAASRDSQV VRREIAKALS KSFTMHSITI
301 QMESPVQDQF DCLFCEDPCD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62f10, frame 2

PIR:S70632 zinc transporter ZnT-2 - rat, N = 1, Score = 884, P = 1.5e-88

TREMBL:MMU76007_1 gene: "ZnT-3"; product: "ZnT-3"; Mus musculus zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 772, P = 1.1e-76

TREMBL:HSU76010_1 gene: "ZnT-3"; product: "ZnT-3"; Human putative zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 742, P = 1.6e-73

TREMBL:MMUZNT02_1 gene: "ZnT-3"; product: "zinc transporter"; Mus musculus zinc transporter (ZnT-3) gene, complete cds., N = 1, Score = 715, P = 1.2e-70

TREMBL:CET18D3_3 gene: "T18D3.3"; Caenorhabditis elegans cosmid T18D3, N = 1, Score = 699, P = 5.9e-69

>PIR:S70632 zinc transporter ZnT-2 - rat
Length = 359

HSPs:

Score = 884 (132.6 bits), Expect = 1.5e-88, P = 1.5e-88
Identities = 171/326 (52%), Positives = 230/326 (70%)

```

Query:      2 YHCHSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID 61
             ++CH+          +E  A+ KL  ASAIC +FMI E++GG++A SLA++TDAHLL D
Sbjct:     34 HYCHAQKDSGSHPNSEKQARRKLYVASAICLVFMIGEIIIGGYLAQSLAINTDAHLLTD 93

Query:      62 LTSFLLSLFSLWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVYLACERLLYPD 121
             S L+SLFSLW+SS+P +K + FGW RAEILGALLS+L IWVVTGVLVYLA +RL+  D
Sbjct:     94 FASMLISLFLWVSSRPATKTMNFGWQRAEILGALLSVLSIWVVTGVLVYLAVQRLISGD 153

Query:     122 YQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH-----KEVQANASVRAAFVHAPG 174
             Y+I+  M+I S CAVA NI++ + LHQ  GH+H          + Q N SVRAAF+H  G
Sbjct:    154 YEIKGDTMLITSGCAVAVNIIMGLALHQSGHGHSHGHSHEDSSQQQONPSVRAAFIHVVG 213

Query:     175 DLFQSIIVLISALIIYFKPEYKIADPICTFIFSIIVLASTITILKDFSILLMEGVPKSLN 234
             DL QS+ VL++A IYFKPEYK DPICTF+FSILVL +T+TIL+D ++LMEG PK ++
Sbjct:    214 DLLQSVGVLVAAAYIYFKPEYKYVDPICTFILSIIVLGTTLTILRDVILVMEGTPKGVGD 273

Query:     235 YSGVKEILILAVDGVLSVHCLHIWSLTMNQVILSAHVATAASRDSQVVRREIAKALSKSFT 294
             ++ VK L+L+VDGV ++H LHIW+LT+ Q +LS H+A A + D+Q V +  L  F
Sbjct:    274 FTTVRNLLLSVDGVEALHSLHIWALTVAQPVLSVHIAIAQNVDAAVLKVARDRLOQGFN 333

```

Pedant information for DKFZphfbr2 62f10, frame 2

Report for DKFZphfbr2 62f10.2

[illegible]

Prosites for DKFZphfbr2 62f10.2

PS000001	162->166	ASN_GLYCOSYLATION	PDOC000001
PS000001	234->238	ASN_GLYCOSYLATION	PDOC000001
PS000004	81->85	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	11->14	PKC_PHOSPHO_SITE	PDOC000005
PS000005	75->78	PKC_PHOSPHO_SITE	PDOC000005

PS00005	80->83	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00007	13->21	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	228->234	MYRISTYL	PDOC00008
PS00013	125->136	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_62f10.2)

DKFZphfbr2_62n10

group: brain derived

DKFZphfbr2_62n10 encodes a novel 541 amino acid protein with similarity to Plasmodium vivax reticulocyte-binding protein 1.

The novel protein contains one Leucine Zipper, involved in protein-protein-interaction. No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to reticulocyte-binding protein

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="13"

Insert length: 3522 bp

Poly A stretch at pos. 3503, polyadenylation signal at pos. 3479

```
1  GGGGCGTGTT  GGC GGGATTC  TGAACGCTGC  CATGGCTCAG  ACCGTGTAGA
51  ATGTTACATT  GTCGCTCACT  CTGCCCATCA  CGTGCCACAT  TTGCTTGGGG
101 AAGGTACGTC  AGCCTGTCAT  ATGCATCAAC  AACCATGTAT  TTTGTTCGAT
151 TTGTATTGAT  TTGTGGTTGA  AGAATAATAG  CCAGTGTCCA  GCTTCAGAG
201 TCCCCATCAC  TCCTGAAAT  CCTTGCAAAG  AAATTATAGG  AGGAACAAGT
251 GAAAGTGAAC  CTATGCTAAG  CCATACGGTC  AGGAAGCATC  TTCGGAAAC
301 TAGACTTGAA  TTACTACACA  AAGAATATGA  GGACGAAATA  GATTGTTTAC
351 AGAAGAAGT  AGAAGAGCTT  AAGAGTAAAA  ATCTCAGCTT  GGAGTCACAG
401 ATCAAAGCTA  TTCTGGATCC  TTTAACCTTG  GTGCAGGGCA  ACCAAAATGA
451 AGACAAACAT  CTAGTCACAG  ATAATCCAAG  TATAATTAAAC  CCAGAACTG
501 TAGCAGAGTG  GAAGAAAAAA  CTCAGAACAG  CTAATGAAAT  CTATGAAAAA
551 GTGAAAGATG  ATGTGGATAA  GCTAAAGGAG  GCAAATAAAA  AATTGAAATT
601 GGAAATATGG  GGTCTGGTGA  GGGAGAATTT  ACGACTGAAG  GCTGAAGTTG
651 ATAACAGATC  ACCTCAAAAG  TTTGGAAGGT  TTGCAGTTGC  TGCTCTTCAG
701 TCCAAAGTAG  AACAGTATGA  GCGTGAAACC  AATCGCCTCA  AGAAAGCCCT
751 GGAACGAAGT  GATAAGTATA  TAGAGGAACT  AGAATCTCAA  GTTGACACGC
801 TAAAAAATTC  AAGTGAAGAG  AAAGAGGCTA  TGAATCCAT  TTGCCAGACA
851 GCACCTTCTG  CAGATGGCAA  AGGGAGCAAA  GGCAGTGAGG  AGGATGTGGT
901 GTCAAAGAA  CAAGGCGATA  GTGCCAGAAA  GCAGCCTGGC  TCATCCACCT
951 CCAGTTCTTC  TCACCTAGCG  AAGCCTTCCA  GCAGCAGACT  GTGTGACACC
1001 AGTTCTGCAA  GGCAGGAAAG  TACCAGCAAA  GCAGACCTTA  ACTGTTCTAA
1051 GAACAAAGAC  CTATATCAAG  AACAGGTAGA  AGTAATGTTA  GATGTGACAG
1101 ATACAAGTAT  GGATACTTAT  TTGGAAAGAG  AATGGGGGAA  TAAACCAAGT
1151 GACTGTGTAC  CCTACAAAGA  TGAAGAACTT  TATGATTTTC  CAGCTCCTTG
1201 TACTCCTTTG  TCCCTTAGTT  GCCTTCAGCT  CAGTACTCCA  GAAATAGAG
1251 AGAGCTCTGT  GGTCCAAGCA  GGAGGTTCCA  AAAAGCACTC  AAACCATCTC
1301 AATGAAATGG  TGTGTGATGA  TTTTGTGAT  TCTTCAAATG  TTTCTAATAA
1351 AGATTCTTCA  GAAGATGATA  TAAGTAGAAG  TGAATATGAG  AAGAAATCAG
1401 AATGTTTTTC  TTCCACAAG  ACAGGATTTT  GGGACTGTTG  TTCCACAAGC
1451 TATGCCCAA  ACTTAGATT  TGAAGTTTCA  GAGGGGAACA  CGATAGCAAA
1501 TTCTGTTGGA  GAAATATCTT  CAAAATTGAG  TGAGAAATCA  GCCTTATGTT
1551 TATCCAAAG  GTTGAATTCT  ATTCGCTCTT  TTGAAATGAA  CCGGACAAGA
1601 ACATCCAGTG  AAGCATCGAT  GGATGCTGCT  TACCTTGACA  AAATCTCTGA
1651 GTTGGATTCA  ATGATGTCAG  AGTCAGACAA  CAGCAAGAGC  CCTTGTAAAT
1701 ACGGTTTTAA  GTCACCTGGT  TTGGATGGGT  TATCAAAGTC  ATCTCAAGGC
1751 AGTGAATTTC  TTGAGGAACC  TGATAAGTTG  GAAGAAAAAA  CTGAGCTAAA
1801 CCTTCCAAA  GGTCTCTTAA  CTAATGATCA  GTTAGAAAT  GGAAGTGAAT
1851 GGAAACCAC  TTCTTTTTTT  TCTCCTCTCT  CCATCTGACC  AAGAAATGAA
1901 TGAAGATTTT  TCACTCCATT  CCAGTTCTTG  TCCAGTAAC  AATGAAATCA
1951 AACCCCAAG  CTGCTTGTTT  CAGACAGAGT  TTTCCAGGG  CATTTTGTTA
2001 AGCAGTTCAC  ATCGACTATT  GGAAGATCAA  AGATTGGGT  CATCTTTGTT
2051 TAAGATGTCC  TCAGAGATGC  ACAGTCTTCA  TAACCACCTT  CAGTCTCCTT
2101 GGTCTACTTC  CTTTGTGCC  GAAAAGAGGA  ATAAAAATGT  GAATCAATCA
2151 ACAAAAAGAA  AAATCCAGAG  CAGCCTTTCC  AGTGCCAGCC  CATCAAAAGC
2201 AACTAAAAGT  TGACTCATT  GAAAGGTGTC  ATTTGTGTT  TTGTCCTGAG
2251 AGAAATAGAA  AAGTTGTTAA  AGTTACCTTT  TTTCTCATA  AAAGTTCTAT
2301 ACAAAATTGA  ATTGATAATC  TTTAGTCAAG  TATCAAGTCA  GGATGGTGGG
2351 TTAACCTGTA  CCCAGAATAC  TTATTGTTCA  TTTGAAAAG  ACTTTGTTCT
2401 TTTCATTTTT  ATTTGGGAGT  CTTTGTGACC  AGAGAAGTTA  GGGAGGAGGT
2451 TATTTTTGTG  TTTTGGGGTT  GGTGGTTGG  TTGGTTTTGT  TTTTGGTTTT
2501 GTTTTTTTAC  TGAATTTGAT  ATGTATCTCG  GTTGATATA  CATTGTTTTT
2551 TTAATAAATG  TTATTTAACT  GTTAGATACA  GTGGCCTGTT  GATAAGCCCC
2601 ACTTGTCTTC  AGAACTTGA  TTTCTTAAT  AAAACTTTTA  GTGTTGCTCA
```

```

2651 TACACTGCTC AATAAGACAC TTGAGTTTAA GCTTTTCCCA GGGTGGAAAT
2701 TATTTTACCT GTCCCTTTTT ATTTATGTTT AGTGATGGCC TAGTTTTTCT
2751 GCAGGGCCAT GATGGAGAAA TAGCACTCTA GCCTTAGTCC AATATTGATT
2801 TACTTTCTTT TTTTAGGTTT TATGTATATG TTTGCATTTT TTAGCATTGT
2851 GTTTTGTCCA GTTTTGTGAA AATGTTCTGC TAGTATGAAA GAAAACATTT
2901 TCTATATGAA GACATTGTGT TTATGTTAGG TAGCTTACAT TTTCTCCTCT
2951 GCGTGTGTGT GTATGTGTGT AAAATCAGAA ATTTAGCATA CTATGGAAAG
3001 AAGGCATGGA GCACTTGGGT TTAGAGGAAC CTAAAACATC ATAGCTTCAT
3051 TGTTCCAGAT GTAACAGGTT TGAAGAGCT CATCGCCAAG TTCTTGATCC
3101 ACTTGCAATC CAGGGGAGTT CTCTTTTGTAG TAGTATGTTT CTTGTTTGCA
3151 TGTTCTGTGT CTTTGTGGAA ACTATGCATG GTAGCATTTT TGCTTGCTGT
3201 GTTTTCCATA CTTAAGAAAA AGAGGTTTCA GTTGGCTGAT AGAATATCTT
3251 TTATGTAGGA CAAAACTTTT CTGTGAAGAG TGTGAGGGG GTGAAGATAG
3301 GTAAGAGGTA AGCACAAATT TTAATTTAGG CTCTGAAAAA GTGTATTGTT
3351 CTAAACGTAT TTGGTATGCC TATATAGGTC TTTAAAAATG GGTTTGTATG
3401 CTGTTTAATG TGCACGAAC ATTTTACATT AATATTGTAC TGTTTTACAT
3451 TAATACTGCA TGCTTTTCTA TGTGAATTGA ATAAAGAATG TCATAAGCAC
3501 TGGAAAAAAA AAAAAAAA AA

```

BLAST Results

Entry HS658254 from database EMBL:
human STS SHGC-11774.
Score = 1643, P = 8.0e-67, identities = 345/355

Entry HS513217 from database EMBL:
human STS SHGC-14656.
Score = 1193, P = 5.8e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 263 bp to 1885 bp; peptide length: 541
Category: similarity to known protein

```

1  MLSHTVRKHL RKTRLELLHK EYEDEIDCLQ KEVEELKSKN LSLESQIKAI
51  LDPLTLVQGN QNEDKHLVTD NPSIINPETV AEWKKKLRTA NEIYKVKDD
101 VDKLKEANKK LKLENGGLVR ENLRLKAEVD NRSPOKEGRF AVAALQSKVE
151 QYERETNRLK KALERSDKYI EELESQVAQL KNSSEEKAM NSICQTALSA
201 DGKSGSGSEE DVVSKNQGDS ARKQPGSSTS SSSHLAKPSS SRLCDTSSAR
251 QESTSKADLN CSKNKDLYQE QVEVMLDVTD TSMDTYLERE WGNKPSDCVP
301 YKDEELYDFP APCTPLSLSC LQLSTPENRE SSVVQAGGSK KHSNHLRKLK
351 FDDFCDSNV SNKDSSEDDI SRSENEKKSE CFSSTKTGFV DCCSTSYAQN
401 LDFESSEGN IANSVGEISS KLSEKSLCL SKRLNSIRSF EMNRTRTSSE
451 ASMDAAYLDK ISELDMMSE SDNSKSPCNN GFKSLDLGL SKSSQGSEFL
501 EEPDKLEET ELNLSKGLT NDQLENGSEW KPTSFPSPLS I

```

BLASTP hits

Entry A42771 from database PIR:
reticulocyte-binding protein 1 - Plasmodium vivax
Score = 127, P = 3.7e-08, identities = 68/300, positives = 145/300

Entry RBP1_PLAVB from database SWISSPROT:
RETICULOCTE BINDING PROTEIN 1 PRECURSOR.
Score = 127, P = 3.9e-08, identities = 68/300, positives = 145/300

Entry MMDSPPG_1 from database TREMBL:
gene: "DSPP"; product: "dentin sialophosphoprotein"; Mus musculus DSPP
gene
Score = 160, P = 5.2e-08, identities = 87/373, positives = 146/373

Alert BLASTP hits for DKFZphfbr2_62n10, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_62n10, frame 2

Report for DKFZphfbr2 62n10.2

```
[LENGTH]      541
[MW]           60533.06
[pI]           5.10
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YKR092c] 3e-05
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKR092c] 3e-05
[PROSITE]      LEUCINE ZIPPER 1
[PROSITE]      MYRISTYL 7
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 18
[PROSITE]      PROKAR_LIPOPROTEIN 1
[PROSITE]      TYR_PHOSPHO_SITE 1
[PROSITE]      PKC_PHOSPHO_SITE 14
[PROSITE]      ASN_GLYCOSYLATION 7
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 9.24 %
[KW]           COILED_COIL 22.55 %
```

```
SEQ      MLSHTVRKHLRKTRLELLHKEYEDEIDCLQKEVEELSKNSLSQIKAILDPLTLVQGN
SEG      .....
PRD      ccechhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheccchhhhheccccccccccc
COILS    .....cccccccccccccccccccccccccccccccccccccccccccc
```

```

SEQ      QNEDKHLVTDNPSIINPETVAEWKKKLRANEIYEKVKDDVDKLEANKKKLLENGGLVR
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      cccceeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceee
COILS    .....ccccccccccccccccccccccccccccccccccccccccccccccccc.....

```

```
SEQ      ENLRLKAEVDNRSPOKFGRFAVAALQSKVEQYERETNLKKALERSDKYTEELESQVAVL
SEG      .....
PRD      ehooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
COILS    ..CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
```

```

SEQ      KNSSEEKAMNSICQTALSADGKSGKSGSEEDVVSKNQGDARSARKQPGSSSTSSSSHLAKPSS
SEG      .....XXXXXXXXXXXXXXXXXXXX
PRD      hcchhhhhhhhhhhhhhhccccccccccceeeccccccccccccccccccccccccccccc
COILS    CCCCCC

```

SEQ SRLCDTSSARQESTSKADLNC SKNKDLYQE QEVMLDVTDTSM DTYLEREWGNKPSDCV P
SEG x.....
PRD ccccccccccccccccccccccc cchhhhhhhccccccccccccchhhhhhhcccccccccc
COILS

SEQ YKDEELYDFPAPCTPLSLSQLSTPENRESSVVQAGGSKKHSNHLRKLVFDDFCDSNNV
 SEG
 PRD cccccccccccccccccceeeccccccccceeecccccccccccccccccccccccccc
 COILS

SEQ SNKDSSEDDISRSENEKKSECFSSTKTGFWDCCSTSQAQLDFESSEGNTIANSVGEISS
 SEG
 PRD cccccccchhhhhhccc
 COILS

SEQ KLSEKSGCLSKRLNSIRSFEMNRTRTSSEASMDAAYLDKISELDSMMSES DNSKSPCNNN
 SEG
 PRD ccccccccchhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccc
 COILS

SEQ GFKSLDGLGSKSSQGSEFLEEPDKLEEKTELNLSKGSLTNDQLENGSEWKPTSFSPLS
SEG . xxxxxxxxxxxxxxxx
PRD ccccccccccccccccecccchhhhhhcccccccccccccccccccccccccc
COILS

SEQ	I
SEG	.
PRD	C
COILS	.

Prosites for DKFZphfbr2 62n10.2

PS00001	40->44	ASN_GLYCOSYLATION	PDOC00001
PS00001	182->186	ASN_GLYCOSYLATION	PDOC00001
PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001

PS00001	359->363	ASN_GLYCOSYLATION	PDOC00001
PS00001	443->447	ASN_GLYCOSYLATION	PDOC00001
PS00001	513->517	ASN_GLYCOSYLATION	PDOC00001
PS00001	526->530	ASN_GLYCOSYLATION	PDOC00001
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	156->159	PKC_PHOSPHO_SITE	PDOC00005
PS00005	166->169	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	361->364	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	419->422	PKC_PHOSPHO_SITE	PDOC00005
PS00005	423->426	PKC_PHOSPHO_SITE	PDOC00005
PS00005	431->434	PKC_PHOSPHO_SITE	PDOC00005
PS00005	436->439	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	183->187	CK2_PHOSPHO_SITE	PDOC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	324->328	CK2_PHOSPHO_SITE	PDOC00006
PS00006	361->365	CK2_PHOSPHO_SITE	PDOC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00006	373->377	CK2_PHOSPHO_SITE	PDOC00006
PS00006	414->418	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	462->466	CK2_PHOSPHO_SITE	PDOC00006
PS00006	469->473	CK2_PHOSPHO_SITE	PDOC00006
PS00007	294->302	TYR_PHOSPHO_SITE	PDOC00007
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	226->232	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	408->414	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	517->523	MYRISTYL	PDOC00008
PS00013	310->321	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_62n10.2)

DKFZphfbr2_62o17

group: metabolism

DKFZphfbr2_62o17.2 encodes a novel 282 amino acid protein with weak similarity to the apolipoprotein E receptor.

The new protein contains a leucine zipper for protein-protein interaction, and three LDL-receptor class A domain (LDLRA_1) patterns. In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands.

The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins

similarity to apolipoprotein E receptor

complete cDNA, complete cds, start at Bp 56 matches kozak consensus
ANCatg EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1260 bp

Poly A stretch at pos. 1240, polyadenylation signal at pos. 1218

```
1  GGGGGATAAG AGAGCGGTCT GGACAGCGCG TGGCCGGCGC CGCTGTGGGG
51 ACAGCATGAG CGGCGGTTGG ATGGCGCAGG TTGGAGCGTG GCGAACAGGG
101 GCTCTGGGCC TGGCGCTGCT GCTGCTGCTC GGCCTCGGAC TAGGCCTGGA
151 GGGCGCCGCG AGCCCGCTTT CCACCCCGAC CTCTGCCCAG GCCCGAGGCC
201 CCAGCTCAGG CTCGTGCCCA CCCACCAAGT TCCAGTGGCG CACCAAGTGGC
251 TTATGCGTGC CCCTCACCTG GCGCTGCGAC AGGGACTTGG ACTGCAGCGA
301 TGCCAGCGAT GAGGAGGAGT GCAGGATTGA GCCATGTACC CAGAAAGGGC
351 AATGCCCCAC GCCCCCTGGC CTCCCTTGCC CCGTGCACCGG CGTCAGTGAC
401 TGCTCTGGGG GAACTGACAA GAAACTGCGC AACTGCAGCC GCCTGGCCCTG
451 CCTAGCAGGC GAGCTCCGTT GCACGCTGAG CGATGACTGC ATTCCACTCA
501 CGTGGCGCTG CGACGGCCAC CCAGACTGTC CCGACTCCAG CGACGAGCTC
551 GGCTGTGGAA CCAATGAGAT CCTCCCGGAA GGGGATGCCA CAACCATGGG
601 GCCCCCTGTG ACCCTGGAGA GCGTCACCTC TCTCAGGAAT GCCACAACCA
651 TGGGGCCCCC TGTGACCCTG GAGAGTGTC CCTCTGTGCG GAATGCCACA
701 TCCTCCTCTG CCGGAGACCA GTCTGGAAGC CCAACTGCCT ATGGGGTTAT
751 TGCAGCTGCT GCGGTGCTCA GTGCAAGCCT GGTCAACCGC ACCCTCCTCC
801 TTTTGTCTCT GCTCCGAGCC CAGGAGCGCC TCCGCCCACT GGGGTTACTG
851 GTGGCCATGA AGGAGTCCCT GCTGCTGTCA GAACAGAAGA CCTCGCTGCC
901 CTGAGGACAA GCACTTGCCA CCACCGTCAC TCAGCCCTGG GCGTAGCCGG
951 ACAGGAGGAG AGCAGTGATG CGGATGGGTA CCCGGGCACA CCAGCCCTCA
1001 GAGACCTGAG CTCCTCTGGC CACGTGGAAC CTCGAACCCG AGCTCCTGCA
1051 GAAGTGGCCC TGGAGATTGA GGGTCCCTGG ACACCTCCCTA TGGAGATCCG
1101 GGGAGCTAGG ATGGGGAACC TGCCACAGCC AGAACCGAGG GGCTGGCCCC
1151 AGGCAGCTCC CAGGGGGTAG GACGGCCCTG TGCTTAAGAC ACTCCTGCTG
1201 CCCCCTCTGA GGGTGGCGAT TAAAGTTGCT TCACATCCTC AAAAAAAAAA
1251 AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 56 bp to 901 bp; peptide length: 282
Category: similarity to known protein
Classification: unset
Prosites motifs: LDLRA_1 (67-90)
LDLRA_1 (67-90)
LDLRA_1 (145-168)


```

1 MSGGWMAQVG AWRTGALGLA LLLLLGLGLG LEAAASPLST PTSQAAGGFS
51 SGSCPPPTKFO CRTSGDSCVLT TWRCDLRDLDC SDGSDEEERC IEPDCTQKQC
101 PPPGLPFCPC TGVSDDCGGPT DKLLRNCRLC ACLAGELRCT LSDDCIPLTW
151 RCDGHDPCCD SSDELGCGTN EILPEGDATT MGPPVTLESV TSLRNATTMG
201 PPVTLESVPS VGNATSSSAG DQSGSTPAYG VTAAAVLVA SLVTATLLLL
251 SWLRAOERLR PLGLLWAMKE SLLLSBKOTS LP

```

No BLASTP hits available

TREMBL:AF110520_6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIF1C, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene., N = 1. Score = 733, P = 1.5e-72

TREMBL:HS275190_1 product: "apolipoprotein E receptor 2 906";
H.sapiens mRNA for apolipoprotein E receptor 2, N = 1, Score = 279, P =
1.8e-23

HSPs:

Query:	6	MAQVGAWRTGALGLALLLLGLGLGLEAAASPLSTPTSAQAAGPSSGSCPPTKFCQRTSG	65
		MA+ GA R ALGL L LL GL GLEAA +P T Q +G + SCP FQC TSG	
Sbjct:	1	MARGGAGRAVALGLVLRLLFLGLRTGLEAAPAHPT--RVQVSGSRADSCPTDTTFCQLTSG	58
Query:	66	LCVPLTWRCDRDLDCSDGSDSEECRIEPICTQKGQCQPPPPGLPCPCTGVSDCSGGTDKKLR	125
		CVPL+WRCD D DCSGSDSEECRIE C Q GQC P LPC C +S CS +DK L	
Sbjct:	59	YCVPLSWRCGDGQDCSDGSDSEEDCRIESCQNGQCQPPQALPSCDNIIGSCSDVSDKNL-	117
Query:	126	NCSRLACLAGELRCTLSDDCIPLTWRCDGHPDCPDSSDELGC GTNEILPEGDATTMGPV	185
		NCSR C EL C L D CIP TWRCDGHPDC DSSDEL C T+	
Sbjct:	118	NCSRPPCQESELHCILDVVICPIHTWRCDGHPDCLDSSDELSCDTD-----T	163
Query:	186	TLESVTSRLRNATTMGPPVTVLESVPSVGNATSSSAGDQSGSP TAYGVIAAAAVLSASLVTA	245
		++ + NATT T+E+ S N T +SAGD S +P+AYGVIAAA VLSA LV+A	
Sbjct:	164	EIDKIFQEENATTTRISTTMENETSFNRVFTFSAGDSSRNPSAYGVIAAAGVLSAILVSA	223
Query:	246	TLLLLSWLRAQERLRPLGLLVAMKESLLLSEQKTS L	281
		TLL+L LR Q L P GLLVA+KESLLLE+KTS L	
Sbjct:	224	TLLIILLRLRGQGYLPPGGLLVAVKESLLLSERKTS L	259

Report for DKFZphfbr2 62o17.2

274

[illegible]

Prosites for DKFZphfbr2 62o17.2

PS01209	67->90	LDLRA_1	PDOC00929
PS01209	67->90	LDLRA_1	PDOC00929
PS01209	145->168	LDLRA_1	PDOC00929
PS00029	17->39	LEUCINE ZIPPER	PDOC00029

Pfam for DKF2phfbr2 62o17.2

HMM_NAME	TNFR/NGFR cysteine-rich region		
HMM	*CpeGtYtD.WNHvpqClpC.trCePEMGQYmvqPCTwTQNT.VC*		
		CP+ ++ + +C+P RC+ ++ +C + ++ +C	
Query	54	CPPTKFCQRTS--GLCVPLTWRCRR--DL----DCSDGSDDEEC	89

```

HMM_NAME      Low-density lipoprotein receptor domain class A
HMM            *tTCeGPDEFQCgSGeMRCIPMsWvCDGDpDCeDWSDEWPENChp*
               C P +FQC+++ C+P+ W+CD D DC D+SDE E+C+
Query         52  GSCP-PTKFCRTSG-LCVPLTWRCRDRLDCSDGSDE--EECRI    91
54.99 (bits) f: 130 t: 169 Target: dkfzphfbr2_62ol7.2 similarity to apolipoprotein E
receptor
Alignment to HMM consensus:
Query         *tTCeGPDEFQCgSGeMRCIPMsWvCDGDpDCeDWSDEWPENChp*
               C + E +C + CIP+ W+CDG PDC D SDE ++C+
dkfzphfbr2    130  LACL-AGELRCTLSD-DCIPLTWRCDGHPDCPDSSDE--LGCGT    169

```

DKF2phfbr2_64a15

 group: nucleic acid management

DKF2phfbr2_64a15 encodes a novel 255 amino acid protein with strong similarity to inorganic pyrophosphatases

Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity.

The new protein can find application as a new enzyme for biotechnologic processes.

strong similarity to inorganic pyrophosphatases

unspliced Intron 212-256 see EST HS1190948

Sequenced by Qiagen

Locus: unknown

Insert length: 1188 bp

Poly A stretch at pos. 1170, polyadenylation signal at pos. 1151

```

1  GGGGGTTGGG  GACCAGTGCA  GGGACCGGGT  CGCGCCGTGC  TATGGCCCTG
51  TACCACACTG  AGGAGCGCGG  CCAGCCCTGC  TCGCAGAATT  ACCGCCTCTT
101 CTTTAAGAAT  GTAACGGGTC  ACTACATTTC  CCCCTTTCAT  GATATTCCTC
151 TGAAGGTGAA  CTCTAAAGAG  GACACTGAGG  CTCAGGCAT  TTTTATAGAC
201 TTGTCTAAGA  TCTGGAAAAT  GGCATTCTTA  TGAAGAAAGC  ACGAAATGAT
251 GAATATGAGA  ATCTGTTTAA  TATGATTGTA  GAAATACCTC  GGTGGACAAA
301 GGCTAAAATG  GAGATTGCCA  CCAAGGAGCC  AATGAATCCC  ATTAAACAAT
351 ATGTAAAGGA  TGGAAAGCTA  CGCTATGTGG  CGAATATCTT  CCCTTACAAG
401 GGTATATAT  GGAATTATGG  TACCTCCCT  CAGACTTGGG  AAGATCCCCA
451 TGAAGAGAT  AAGAGCACGA  ACTGCTTTGG  AGATAATGAT  CCTATTGATG
501 TTTGCGAAAT  AGGCTCAAAG  ATTCTTTCTT  GTGGAGAAGT  TATTCATGTG
551 AAGATCCTTG  GAATTTTGGC  TCTTATTGAT  GAAGGTGAAA  CAGATTGGAA
601 ATTAATTGCT  ATCAATGCGA  ATGATCCTGA  AGCCTCAAAG  TTTTCATGTA
651 TTGATGATGT  TAAGAAAGTC  AAACCGGGTT  ACCTGGAAGC  TACTCTTAAT
701 TGGTTTAGAT  TATGTAAGGT  ACCAGATGGA  AAACCGAGAA  ACCAGTTTGC
751 TTTTAATGGA  GAATTCAAAA  ACAAGGCTTT  TGCTCTTGAA  GTTATTAAAT
801 CCACTCATCA  ATGTTGGAAA  GCATTGCTTA  TGAAGAACTG  TAATGGAGGA
851 GCTACAAATT  GCACAAACGT  GCAGATATCT  GATAGCCCTT  TCCGTTGCAC
901 TCAAGAGGAA  GCAAGATCAT  TAGTTGAATC  GGTATCATCT  TCACCAAATA
951 AAGAAAGTAA  TGAAGAAGAG  CAAGTGTGGC  ACTTCCTTGG  CAAGTGATTG
1001 AAACATCTGA  AATTCTGCTG  TCAAGATTCC  CATCTCTAAG  GACTCCAAGA
1051 CTCCTTTTCC  CCAAGTGCTA  GAGACAAGGG  GGTCTATGAG  CATTCTACTG
1101 CTTCTGTGTA  AAACCTCATT  TTTTCAAAC  TTTTGAGCTA  TGCAATATAT
1151 AAATAAACAG  TAAGAATTTT  AAAAAAAAA  AAAAAAAAA

```

BLAST Results

Entry HSPPAEMR from database EMBL:
 H.sapiens partial mRNA for pyrophosphatase.
 Score = 1706, P = 1.6e-70, identities = 342/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 230 bp to 994 bp; peptide length: 255
 Category: strong similarity to known protein
 Classification: unset
 Prosite motifs: PPASE (85-92)

```

1 MKKARNDEYE NLFNMIVEIP RWTAKMEIA TKEPMNPIKQ YVKDGKLRVY
51 ANIFPYKGYI WNYGTLFQTW EDPHEKDKST NCFGDNDPID VCEIGSKILS
101 CGEVIHVKIL GILALIDEGE TDWKLIANA NDPEASKFHD IDDVKKFKPG
151 YLEATLNWFR LCKVPDGKPE NQFAFNGEFK NKAFALVNIK STHQCWKALL
201 MKNCNGGATN CTNVQISDSP FRCTQEEARS LVESVSSSPN KESNEEQVW
251 HFLGK

```

BLASTP hits

Entry IPYR_KLULA from database SWISSPROT:
INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE PHOSPHO-
HYDROLASE) (PPASE).
Score = 689, P = 6.0e-68, identities = 128/248, positives = 170/248

Entry A45153 from database PIR:
inorganic pyrophosphatase (EC 3.6.1.1) - bovine
Score = 862, P = 2.8e-86, identities = 146/226, positives = 190/226

Entry AF085600.1 from database TREMBLNEW:
gene: "Nurf-38"; product: "inorganic pyrophosphatase NURF-38";
Drosophila melanogaster inorganic pyrophosphatase NURF-38 (Nurf-38)
gene, complete cds.
Score = 731, P = 2.1e-72, identities = 134/248, positives = 177/248

Entry PWBV from database PIR:
inorganic pyrophosphatase (EC 3.6.1.1) - yeast (Saccharomyces
cerevisiae)
Score = 688, P = 7.7e-68, identities = 133/251, positives = 174/251

Alert BLASTP hits for DKFZphfbr2_64a15, frame 2

SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
(PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 731, P =
2.4e-72

>SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
PHOSPHO- HYDROLASE) (PPASE).
Length = 290

HSPs:

Score = 731 (109.7 bits), Expect = 2.4e-72, P = 2.4e-72
Identities = 134/248 (54%), Positives = 177/248 (71%)

```

Query:      7 DEYENLFNMIVEIPRWTAKMEIATKEPMNPIKQYVKDGKLRVYANIFPYKGYIWNNGTL 66
             +E + ++NM+VE+PRWT AKMEI+ K PMNPIKQ +K GKLR+VAN FP+KGYIWNNG L
Sbjct:     40 NEEKTIYNMVVEVPRWTNAKMEISLKTMPNPIKQDIKKGKLRVANCFFPHKGYIWNNGAL 99

Query:     67 PQTWEDPHEKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGETDWKLI 126
             PQTWE+P + ST C GDNDPIDV EIG ++ G+V+ VK+LG ALIDEGETDWK+I
Sbjct:    100 PQTWENPDHIEPSTGCKGNDNDPIDVIEIGYRVAKRGDVLKVKVLGQFALIDEGETDWKII 159

Query:    127 AINANDPEASKFHDIDDVKKFKPGYLEATLNWFR LCKVPDGKPENQFAFNGEFKNKAFAL 186
             AI+ NDP ASK +DI DV ++ PG L AT+ WF++ K+PDGKPENQFAFNG+ KN FA
Sbjct:    160 AIDVNDPLASKVNDIADVQYFPGLLRATVEWFKIYKIPDGKPENQFAFNGDAKNADFAN 219

Query:    187 EVIKSTHQCWKALLMKNCNGGATNCTNVQISDSPFRCTQEEARS-LVESVSSSPNKESNE 245
             +I TH+ W+ L+ ++ G+ + TN+ +S +EEA L E+ +E ++
Sbjct:    220 TIIAETHKFWQNLVHQSPASGSISTTNITNRNSEHVIPKEEA EKILAEAPDGGQVEEVSD 279

Query:    246 EEQVWHFL 253
             WHF+
Sbjct:    280 TVDTWHFI 287

```

Peptide information for frame 3

ORF from 42 bp to 230 bp; peptide length: 63
Category: strong similarity to known protein
Classification: unset

```

1 MALYHTEERG QPCSQNYRLF FKNVTGHYIS PFHDIPLKVN SKEDTEAQGI
51 FIDLSKIWKM AFL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64a15, frame 3

SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
(PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 118, P = 8.8e-07

PIR:A45153 inorganic pyrophosphatase (EC 3.6.1.1) - bovine, N = 1,
Score = 113, P = 3.1e-06

TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase";
Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds., N
= 1, Score = 106, P = 1.8e-05

>SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
PHOSPHO- HYDROLASE) (PPASE).
Length = 290

HSPs:

Score = 118 (17.7 bits), Expect = 8.8e-07, P = 8.8e-07
Identities = 23/43 (53%), Positives = 29/43 (67%)

Query: 1 MALYHTEERGQPCSONYRLFFKNVTGHYISPFHDIPLKVNSKE 43
MALY T E+G S +Y L+FKN G+ ISP HDIPL N ++
Sbjct: 1 MALYETVEKGAKNSPSYSLYFKNKGNVISPMHDIPLYANEK 43

Pedant information for DKFZphfbr2_64a15, frame 2

Report for DKFZphfbr2_64a15.2

[LENGTH] 255
[MW] 29177.34
[pI] 5.67
[HOMOL] TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase"; Homo
sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds. 2e-93
[FUNCAT] 01.04.01 phosphate utilization [S. cerevisiae, YBR011c] 9e-73
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR011c] 9e-73
[FUNCAT] 02.99 other energy generation activities [S. cerevisiae, YMR267w] 1e-58
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR267w] 1e-58
[FUNCAT] 1 genome replication, transcription, recombination and repair [M.
genitalium, MG351] 1e-06
[FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI0124] 2e-06
[BLOCKS] BL00387D
[BLOCKS] BL00387C
[BLOCKS] BL00387B
[BLOCKS] BL00387A
[SCOP] dlwgja_2.29.5.1.1 Inorganic pyrophosphatase {baker's yeas 1e-113
[EC] 3.6.1.1 Inorganic pyrophosphatase 7e-92
[PIRKW] mitochondrion 3e-57
[PIRKW] hydrolase 7e-92
[PIRKW] homodimer 2e-71
[SUPFAM] inorganic pyrophosphatase 7e-92
[PROSITE] PPASE 1
[KW] Alpha_Beta
[KW] 3D
[KW] LOW_COMPLEXITY 6.27 %

SEQ MKKARND EYENLFNMIVEIPRWT KAKMEIATKEPMNPIKQYVKDGLRYVANIFPYKGYI
SEG
lhukBEGGCCEEEEEETTTbCBCEETTTTTTCEEECEETTEECBCCBTTBTtbt

SEQ WNYGTLPTQWEDPHEKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGE
SEG
lhukB CEEETTTTCBTTTTEETTTTEECCECCBCEEEECCECCCTTTEEEEEEEEEETTTTB

SEQ TDWKLIAINANDPEASKFHDIDVKKFKPGYLEATLNWFLCKVPDGKPENQFAFNGEFK
SEG
lhukB CEEEEEEETTTTGGGCCCHHHHHHTTTHHHHHHHHHHHC GGGCCCCCBGGGCCB

SEQ NKAFALEVIKSTHCWKALLMKNCNGGATNCTNVQISDSPFRCTQEARSLVESVSSSPN
SEGxxxxxxx
lhukB CHHHHHHHHHHHHHHHHHCTTTTTTCCCBTTTTTTT.....

DKFZphfbr2_64c16

group: brain derived

DKFZphfbr2_64a16.2 encodes a novel 101 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: /map="745_A_2; 756_F_2; 842_C_2"

Insert length: 1866 bp

Poly A stretch at pos. 1848, polyadenylation signal at pos. 1829

```
1 GGGCGCGGCG CCGGAGGAGG AAGTGGTGAG GTTGTGCTC CTTACGCGCC
51 TATCGCTGGC TCTTGGGGCG CAGAGAGGGG CCGCAGTCTC CGCGGCTGCG
101 TCGAGCTCCC TTGCAGTCCC CTCCATGTTC CCGGCGGCCA CTACTCCCT
151 TCCTAAGGCC GCCGCTTACC CCGGGGTCTA TGAAGTAAT GGAAGGACCC
201 CTCACCTGG CTCATCAACA GAGCAGACGA GCAGACCGTT TATTAGCTGC
251 AGGCAAATAC GAAGAGGCTA TTTCTTGTC CAAAAGGCT GCAGCATATC
301 TTTCTGAAGC CATGAAGCTG ACACAGTCAG AGCAGGCTCA TCTTCTACTG
351 GAATTGCAAA GGGATAGCCA TATGAAACAG CTCCTCCTCA TCCAAGAGAG
401 ATGGAAAAGG GCCCAGCGTG AAGAAAGATT GAAAGCCCAG CAGAACACAG
451 ACAAGGATGC AGCTGCCCAT CTTCAGACAT CTCACAAACC CTCTGCAGAG
501 GATGCAGAGG GCCCAGAGTCC CCTTCTCTAG AAGTACAGCC CTCCACAGA
551 GAAATGCCTG CCTGAGATTG AGGGGATCTT TGACAGGGAT CCAGACACAC
601 TACTTTATTT ACTTCAGCAA AAGAGTGAGC CAGCAGAGCC ATGTATTGGA
651 AGCAAAGCCC CAAAAGATGA TAAAACAATT ATAGAGGAGC AGGCAACCAA
701 AATTGTCAGT TTGAAGAGGC ATGTGGAATT CCTTGTGGCT GAGAATGAAA
751 GATTAAGGAA AGAAAATAAA CAACTAAAGG CTGAAAAGGC CAGACTTCTA
801 AAAGGTCCAA TAGAAAAGGA GCTGGATGTA GATGCTGATT TTGTAGAAAC
851 GTCAGAGTTA TGGAGCTTGC CACCACATGC AGAACTGCT ACAGCCTCCT
901 CAACCTGGCA GAAGTTCGCA GCAAATACTG GGAAAGCCAA GGACATTCCA
951 ATCCCAATC TTCTCCCTT GGATTTTCCA TCTCCAGAAC TTCCTCTTAT
1001 GGAGCTCTCT GAGGATATTC TGAAAGGACT TATGAATAAT TAAAATGGAA
1051 GGCCACAGAA AAGGGGAAAA GAGGAAATAA TACAGTAATC GTTAATCCAG
1101 CAAAAAGAAA TGAAAAGGGA AAACACATA GAAGGGTAAT CCCGGAATG
1151 CTTCATCTGG TGGACTGTGG GAGCAGAGGC ATTGCCAGGA CTTGGGAAAC
1201 AGTCACTGTG AAATGCGCTG CGTATCTCAT TCACTCACTT CAGCTAATGA
1251 CTCCGACTTG GCAGACGCTA AACTCATGGA GGTTCGTTT CTCCTGATAC
1301 AAACCAAATG GCTACCTGGA AGAATTTCTT TCAAGCAACA GTTATTTTTC
1351 TTATCTTCAG GGTAAAAATG TATAAAGTT ATGTGTAATT AATCTATAAT
1401 GCCATAAATG ATAATGCAAA ACCTAAATAA TATGGTGGCC GGAGGGGCTG
1451 CCTTATATTT GAAACATGCT TTCTATCATG CATTGACTGT ATGCATTTTG
1501 TTAATGCACA TTCTGTTTGT TTAAGGTGTG TGAGATACAC ACCTTCTAG
1551 ATGAAACTAT ATGTGCCACA CTTGCACTA CTCATAATGA TAACCTCAAG
1601 ACTATCAGAA GAAATATTTA AATTTCCATT TTATGAAGAA AGGAACCAA
1651 TTATTATGCT TTTTAAACA AATTACCAGT TTACATAATT AATCAGGGTG
1701 CATTTTAAAG TCTAACTTCG TTTATTGTAT AATGCATCAT TTGAAAATAC
1751 CAAGGAGGAA ATACCTTTTG TTTTAAATGA TGCAAGAGTG GACGTAATGC
1801 TAGTTGGCAG TATTTTATTG TAAGAAATCA ATAAAGTAAT TGTGTTTAA
1851 AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS286143 from database EMBL:
human STS WI-6844.
Score = 1460, P = 3.4e-61, identities = 292/292

Medline entries

No Medline entry

.....

1 GAAPEEEVVR LLLLQRLSLA LGAQRGA AVS AAASSSLAVP SMFPGATTPL
51 PKAAAYPGVY GSNGRTPQPG SSTEQTSRPF ISCRQIRRGY FLSQKGCSIS
101 F

No BLASTP hits available

No Alert BLASTP hits found

1	MEVMEGPLNL	AHQSSRRADR	LLAAGKYEEA	ISCHKKAAY	LSEAMKLTQS
51	EQAHLSLEAQ	RDSHMQQLL	IQERWKRAR	EEERLKAQNT	DKDAAHLQT
101	SHKPSADEAE	GQSPSQSQS	PSTEKCLPEI	QGIFDRPDPT	LLYLLQKSE
151	PAEPCIGSKA	PKDDKTIIE	QATKIADLR	HVEFLVAENE	RLRKENKQLK
201	AEAKARLLKP	IEPLDVDPD	FVETSELSDI	PPHATATAS	STWQKFAANT
251	GKAKDIPNP	LEPLDFSPSE	LPMLSELSDI	LKGLMNN	

No BLASTP hits available

No Alert BLASTP hits found

```
[LENGTH]          101
[MW]               10469.94
[pI]               10.18
[KW]               All_Alpha
[KW]               LOW_COMPLEXITY      29.70 %
```

(No Prosite data available for DKFZphfbr2 64c16.2)

(No Pfam data available for DKFZphfbr2_64c16.2)

Report for DKFZphfbr2_64c16.3

```

{LENGTH}      287
{MW}           32343.79
{PI}           5.61
{PROSITE}      LEUCINE_ZIPPER 2
{KW}           All Alpha
{KW}           COILED_COIL      14.98 %

SEQ  MEVMEGPLNLAHQQSRRADRLAAGKYEEAISCHKKAAAYLSEAMKLTQSEQAHLSELEQ
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ  RDSHMQLLLIQRWKRQREERLKAQQNTDKDAAHLQTSCHKPSAEDAEGQSPLSQKYS
PRD  hhccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ  PSTEKCLPEIQGIFDRDPDTLLYLLQKSEPAEPCIGSKAPKDDKTIIEEQATKIADLKR
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....CCCCCCCCCCCCC

SEQ  HVEFLVAENERLRKENKQLKAEKARLLKGPIEKELDVDADFVETSELWSLPPHAETATAS
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ  STWQKFAANTGKAKDIPINLPPLDFPSPPELPLMELSEDILKGLMNN
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```

Prosites for DKFZphfbr2_64c16.3

```

PS00029  178->200  LEUCINE_ZIPPER      PDOC00029
PS00029  185->207  LEUCINE_ZIPPER      PDOC00029

```

(No Pfam data available for DKFZphfbr2_64c16.3)

DKFZphfbr2_64c4

group: brain derived

DKFZphfbr2_64c4 encodes a novel 467 amino acid protein with similarity to A. thaliana T08I13.5

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A. thaliana T08I13.5

complete cDNA, complete cds, EST hits
on genomic level encoded by AC005043 11 exons

Sequenced by Qiagen

Locus: unknown

Insert length: 1559 bp

Poly A stretch at pos. 1540, no polyadenylation signal found

```
1 TGGGACCGCC GGAAGTTTCT GCCGCGGCTT TGCAGGGGACG GGGGAGTGGT
51 AGTGGGGGCT GCAGCTGCCG GACCCAGGCG CGATGGCTAC GGGCGCGGAT
101 GTACGGGACA TTCTAGAACT CGGGGGTCCA GAAGGGGATG CAGCCTCTGG
151 GACCATCAGC AAGAAGGACA TTATCAACCC GGACAAGAAA AAATCCAAGA
201 AGTCCTCTGA GACACTGACT TTCAAGAGGC CCGAGGGCAT GCACCGGGAA
251 GTCTATGCCT TGCTCTACTC TGACAAGAAG GATGCACCCC CACTGCTACC
301 CAGTGACACT GGCCAGGGAT ACCGTACAGT GAAGGCCAAG TTGGGCTCCA
351 AGAAGGTGCG GCCTTGGAAG TGGATGCCAT TCACCAACCC GGGCCGCAAG
401 GACGGAGCAA TGTCTTTCCA CTGGCGACGT GCAGCGGAGG AGGGCAAGGA
451 CTACCCCTTT GCCAGGTCCA ATAAGACTGT GCAGGAGCCT GTGTACTCGG
501 AGCAGGAGTA CCAGCTTTAT CTCCACGATA ATGCTTGGAC TAAGGCAGAA
551 ACTGACCACC TCTTTGACCT CAGCCGCCGC TTTGACCTGC GTTTTGTGTG
601 TATCCATGAC CGGTATGACC ACCAGCAGTT CAAGAAGCGT TCTGTGGAAG
651 ACCTGAAGGA GCGGTACTAC CACATCTGTG CTAAGCTTGC CAACGTGCGG
701 GCTGTGCCAG GCACAGACCT TAAGATACCA GTATTTGATG CTGGGCACGA
751 ACGACGGCGG AAGGAACAGC TTGAGCGTCT CTACAACCGG ACCCCAGAGC
801 AGGTGGCAGA GGAGGAGTAC CTGCTACAGG AGCTGCCCAA GATTGAGGCC
851 CGAAGAAGAG AGCGGGAGAA ACGCAGCCAG GACCTGCAGA AGCTGATCAC
901 AGCGGCAGAC ACCACTGCAG AGCAGCGCGC CACGGAACGC AAGGCCCCCA
951 AAAAGAAGCT ACCCCAGAAA AAGGAGGCTG AGAAGCCGGC TGTTCCTGAG
1001 ACTGCAGGCA TCAAGTTTCC AGACTTCAAG TCTGCAGGTG TCACGCTGCG
1051 GAGCCAACGG ATGAAGCTGC CAAGCTCTGT GGGACAGAAG AAGATCAAGG
1101 CCTTGGAAAC GATGCTGCTG GAGCTTGGTG TGGAGCTGAG CCCGACACCT
1151 ACGGAGGAGC TGGTGCACAT GTTCAATGAG CTGCGAAGCG ACCTGGTGCT
1201 GCTCTACGAG CTCAAGCAGG CCTGTGCCAA CTGCGAGTAT GAGCTGCAGA
1251 TGCTGCGGCA CCGTCATGAG GCACTGGCCC GGGCTGGTGT CTAGGGGGGC
1301 CTTGCCACAC CAGCATCAGG CCCAGGCCCG GCCTCTGCTG AGCCGGCAGT
1351 GTCTGAACCC GGAATTGGTC CTGACCCCAA GGACACCATC ATTGATGTGG
1401 TGGGCGCACC CCTCACGCC AATTGAGAA AGCGACGGGA GTCGGCCTCC
1451 AGTCTATCTT CCGTGAAGAA AGCCAAGAA CCGTGAGAGG CCCCACGGGG
1501 TGTGGCGGAC GCTGTATGT AAATAGAGCT GCTGAGTTGG AAAAAAAAAA
1551 AAAAAAAAAA
```

BLAST Results

Entry AC005043 from database EMBL:
Homo sapiens clone NH0576N21; HTGS phase 1, 5 unordered pieces.
Score = 1506, P = 4.6e-244, identities = 316/330

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 83 bp to 1483 bp; peptide length: 467

Category: similarity to unknown protein

```
1 MATGADVVDI LELGGPEGDA ASGTISKDDI INPDKKKSKK SSETLTFRKP
51 EGMHREYVAL LYSDKKDAPP LLPSDTGQGY RTVKAKLGSK KVRPWKWMPP
101 TNPARKDGAM FFHWRRAAEE GKDYFFARFN KTVQEPVYSE QEYQLYLHDN
151 AWTKAETDHL FDLRRFDLR FVVIHRYDH QQFKKRSVED LKERYHYICA
201 KLANVRAVPG TDLKIPVFDA GHERRRKEQL ERLYNRTPEQ VAEEEYLLQE
251 LRKIEARKKE REKRSQDLQK LITAADTTAE QRRTERKAPK KKLPPQKEAE
301 KPAVPETAGI KFPDFKSAGV TLRQRMKLP SSVGQKKIKA LEQMLLELGV
351 ELSPTPTEEL VHMFNELRSD LVLLYELKQA CANCEYELQM LRHRHEALAR
401 AGVLGGPATP ASGPGPASAE PAVSEPLGP DPKDTIIDVV GAPLTPNSRK
451 RRESASSSSS VKKAKKP
```

BLASTP hits

Entry ATAC2337_5 from database TREMBLNEW:
gene: "T08I13.5"; Arabidopsis thaliana chromosome II BAC T08I13
genomic sequence, complete sequence.
Score = 340, P = 2.6e-30, identities = 115/374, positives = 176/374

Entry YE8D_SCHPO from database SWISSPROT:
HYPOTHETICAL 47.1 KD PROTEIN C9G1.13C IN CHROMOSOME I.
Score = 221, P = 1.9e-20, identities = 67/192, positives = 97/192

Entry S64291 from database PIR:
hypothetical protein YGR002c - yeast (*Saccharomyces cerevisiae*)
Score = 202, P = 2.8e-13, identities = 71/260, positives = 124/260

Alert BLASTP hits for DKFZphfbr2_64c4, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64c4, frame 2

Report for DKFZphfbr2_64c4.2

```
[LENGTH]      467
[MW]           53007.60
[pI]           9.51
[HOMOL]        TREMBL:ATAC2337_5 gene: "T08I13.5"; Arabidopsis thaliana chromosome II BAC
T08I13 genomic sequence, complete sequence. 4e-29
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGR002c] 1e-19
[PROSITE]      MYRISTYL 1
[PROSITE]      CAMP_PHOSPHO_SITE 4
[PROSITE]      CK2_PHOSPHO_SITE 10
[PROSITE]      TYR_PHOSPHO_SITE 3
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 12
[PROSITE]      ASN_GLYCOSYLATION 1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 20.13 %
```

```
SEQ  MATGADVVDILELGGPEGDAASGTISKDDIINPDKKKSKKSSETLTFRKP EGMHREYVAL
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhh
```

```
SEQ  LYSDKKDAPPLPSDTGQGYRTVKAKLGSKKVRPWKWMPPFTNPARKDGAMFFHWRRAAEE
SEG  .....
PRD  hhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhh
```

```
SEQ  GKDYFFARFNKTVQEPVYSEQEYQLYLHDNAWTKAETDHLFDLSRRFDLRFVVIHRYDH
SEG  .....
PRD  cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeecccc
```

```
SEQ  QQFKKRSVEDLKERYHYICAKLANVRAVPGTDLKIPVFDAGHERRRKEQLERLYNRTPEQ
SEG  .....
PRD  chhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

```
SEQ  VAEEEYLLQELRKIEARKKEREKRSQDLQKLITAADTTAEQRRTERKAPKKLPPQKEAE
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

```
SEQ  KPAVPETAGIKFPDFKSAGVTLRSQRMKLPSSVGQKKIKALEQMLLELGVLSPTPTEEL
SEG  xxx.....
```

```

PRD      hccccccccccccccccceehhhhhhhccccccchhhhhhhhhhhhhhhhhhhhhccccchhh
SEQ      VHMFNELRSDLVLLYELKQACANCEYELQMLRHRHEALARAGVLGGPATPASGPGPASAE
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhccccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ      PAVSEPLGPDPKDTIIDVVGAPLTPNSRKRRESASSSSSVKKAKKP
SEG      xxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxx.
PRD      cccccccccccccceeecccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_64c4.2

PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00002	412->416	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	35->39	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	39->43	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	184->188	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	451->455	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	284->287	PKC_PHOSPHO_SITE	PDOC00005
PS00005	321->324	PKC_PHOSPHO_SITE	PDOC00005
PS00005	324->327	PKC_PHOSPHO_SITE	PDOC00005
PS00005	448->451	PKC_PHOSPHO_SITE	PDOC00005
PS00005	460->463	PKC_PHOSPHO_SITE	PDOC00005
PS00006	3->7	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	153->157	CK2_PHOSPHO_SITE	PDOC00006
PS00006	187->191	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	435->439	CK2_PHOSPHO_SITE	PDOC00006
PS00007	131->139	TYR_PHOSPHO_SITE	PDOC00007
PS00007	227->235	TYR_PHOSPHO_SITE	PDOC00007
PS00007	116->125	TYR_PHOSPHO_SITE	PDOC00007
PS00008	14->20	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_64c4.2)

DKFZphfbr2_64h6

group: brain derived

DKFZphfbr2_64h6 encodes a novel 176 amino acid protein with similarity to predicted yeast proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to S.pombe SPBC337.09 and S.cerevisiae YER044c

complete cDNA, complete cds accoring to YER044c/SPBC337.09,
start at Bp 111, EST hits

Sequenced by Qiagen

Locus: /map="14"

Insert length: 1212 bp

Poly A stretch at pos. 1192, polyadenylation signal at pos. 1168

```
1 GGGCTGGAGC TGTCCTGGGG GAGCTTGTTT GCGGCAGCGG CTGCTGCTGC
51 CACTGCTGTG CTGGGGGGCC GGTCGCCAGG CAAAAAGCCC TCCCACGTTT
101 GAGGGGAGTC ATGAGCCGTT TCCTGAATGT GTTAAGAAGT TGGCTGGTTA
151 TGGTGTCAT CATAGCCATG GGGAACACGC TGCAGAGCTT CCGAGACCAC
201 ACTTTTCTCT ATGAAAAGCT CTACACTGGC AAGCCAAACC TTGTGAATGG
251 CCTCCAAGCT CGGACCTTTG GGATCTGGAC GCTGCTCTCA TCAGTGATCC
301 GCTGCCTCTG TGCCATTGAC ATTCACAACA AGACGCTCTA TCACATCACA
351 CTCTGGACCT TCCTCCTTGC CCTGGGGCAT TTCTCTCTG AGTTGTTTGT
401 CTATGGAACG GCAGCTCCCA CGATTGGCGT CCTGGCACCC CTGATGGTGG
451 CAAGTTTCTC CATCTGGGT ATGCTGGTCG GGCTCCGGTA TCTAGAAGTA
501 GAACCAAGTAT CCAGACAGAA GAAGAGAAAC TGAGGCCAGC ATTATCACCT
551 CCAGGACTTT CTCGTTTCC ACCTTGGCCA TCTTCTTCTC TCGTCGTCTC
601 TCCCTTTTAA TTTCTTTTCT ATTCCATCAT CTGCCCTTTT ACTCACTTTT
651 AGCCTCTTTT TTTAATTTT AAAATTAAA GATATGCATA CTGAAAAGTA
701 TATAACATGT ACGTACAATT TAAAGAATAA TTTTAAAGTG AATACTACGT
751 AACTCCATCC AAGTCAAGAA ATTGCCAGCT TCTCGGAAGC CCACTGTGTC
801 TCCTTCCCTT ACCTGCAACC TCTTCCAGGC TCCCTTTTCC AGCCTTCCCC
851 TTTTCCCTT TTTATTTTAT GCCTTGATTT GACTTGTGTG GTGGGAACAT
901 GTGAACTATG AAACCTTAAAC CTGCTGCCCA CCCAGAGCAG CTGTGACCAA
951 GGGCTGCCTC AAGGGGTTGT CCACGCAGGT TGGGCTCCTC TCTGCTGCTG
1001 GACCCAAGAC TCTGAACCTT CCAAGGGACA GGCAGTTCTT CTGAGAAGGG
1051 CTCCCCTGTG TGTAGCAAG ACCACAGCTC TCCTTCTATC TACAGATGCA
1101 TGAGGGTGG AAGAGTCTGG GCTGTTTTTA GACCTTCTGG TCAGCTGTAT
1151 TTGTGTAACA ACTTTGTAA TAAATAGAAA AACCTCTGCT TCAAAAAAAA
1201 AAAAAAAA AA
```

BLAST Results

Entry G38566 from database EMBL:

SHGC-64295 Human Homo sapiens STS genomic, sequence tagged site.

Score = 1398, P = 1.4e-56, identities = 284/288

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 0 bp to 530 bp; peptide length: 177

Category: similarity to unknown protein

Classification: unclassified

```
1 AGAVLGELVC GSGCCCHCCA GGPVARQKAL PRLRGVMSRF LNVLRSLVLM
51 VSIIAMGNFL QSRDHTFLY EKLYTGKPNL VNGLQARTFG IWTLLSSVIR
101 CLCAIDIHKN TLYHITLWTF LLALGHFLSE LFVYGTAAPT IGVLAFLMVA
```

151 SFSILGMLVG LRYLEVEPVSRQKKRN

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64h6, frame 3

TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein"; S.pombe chromosome II cosmid c337., N = 1, Score = 224, P = 1.4e-18

PIR:S50547 hypothetical protein YER044c - yeast (Saccharomyces cerevisiae), N = 1, Score = 192, P = 3.4e-15

>TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein"; S.pombe chromosome II cosmid c337.
Length = 136

HSPs:

Score = 224 (33.6 bits), Expect = 1.4e-18, P = 1.4e-18
Identities = 49/113 (43%), Positives = 74/113 (65%)

Query: 42 NVLRSLVLMVSIAMGNTLQSFDRDHTFLYKLYTGKPNLVNGLQARTFGIWTLLSSVIRC 101
+++ W V+VS+ A+ NT+QSF L +++Y+ N VNGLQ RTFGIWTLLS+++R
Sbjct: 11 SLVAKWNVVSVAAALFNTVQSFLTFR-LTKRVYSNT-NEVNGLQGRTFGIWTLLSAIVRF 68
Query: 102 LCAIDIHNKTLHYHITLWTFLLALGHFLSELFVYGTAAPTIGVLAPLMVASFSI 154
CA I N +Y + T+ LA HFLSE ++ T G+L+P++V++ SI
Sbjct: 69 YCAYHITNPDVYFLCQCTYYLACFHFLEWLLFRTTNLGPGLLSPIVVSTVSI 121

Pedant information for DKFZphfbr2_64h6, frame 3

Report for DKFZphfbr2_64h6.3

[LENGTH] 176
[MW] 19359.31
[pI] 9.53
[HOMOL] TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein";
S.pombe chromosome II cosmid c337. 2e-17
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YER044c] 7e-16
[KW] TRANSMEMBRANE 2
[KW] LOW_COMPLEXITY 7.39 %

SEQ AGAVLGELVCGSGCCCHCCAGGPVARQKALPRLRGVMSRFLNVLRSLVLMVSIAMGNTL
SEGxxxxxxxxxxxxx.....
PRD cccceeeeeeeccceeeccccccccccccccccchhhhhhhhhhhhhheeeccccc
MEMMMMMMMMMMMMMMMMM.....

SEQ QSFDRDHTFLYKLYTGKPNLVNGLQARTFGIWTLLSSVIRCLCAIDIHNKTLHYHITLWTF
SEG
PRD cccccchhhhhhhhhccccccccccccccccchhhhhhhhhhhhhccccceeehhhhh
MEM
SEQ LLALGHFLSELFVYGTAAPTIGVLAPLMVASFSILGMLVGLRYLEVEPVSRQKKRN
SEG
PRD hhhhhhhhhhhhhccccccccceeehhhhhhhhhhheeecccccccccc
MEMMMMMMMMMMMMMMMMM.....

(No Prosite data available for DKFZphfbr2_64h6.3)

(No Pfam data available for DKFZphfbr2_64h6.3)

DKFZphfbr2_64j18

group: Intracellular transport and trafficking

DKFZphfbr2_624j18.1 encodes a novel 180 amino acid protein nearly identical to the microsomal signal peptidase 23 kd subunit of canis familiaris, gallus gallus and C. elegans.

The new protein is identical to canine and chicken microsomal signal peptidase 23 kd subunit. The canine microsomal signal peptidase is a protein complex comprised of five subunits (25, 22/23, 21, 18, and 12 kDa). The 23kDa subunit is tightly associated with the 18- and 21-kDa subunits, that are integral membrane proteins.

The new protein can find application in modulation of protein transport into microsomal compartments and as a tool for proteomic analysis.

strong similarity to dog signal peptidase (EC 3.4.99.-)

complete cDNA, complete cds, potential start at Bp 109, EST hits,

Sequenced by Qiagen

Locus: unknown

Insert length: 690 bp

Poly A stretch at pos. 666, polyadenylation signal at pos. 646

```

1 GCCGGAACGC GCGCACCGCA GACGGCGCGG ATCGCAGGGA GCCGGTCCGC
51 CGCCGGAACG GGAGCCTGGG TGTGCGTGTG GAGTCCGGAC TCGTGGGAGA
101 CGATCGCGAT GAACACGGTG CTGTCGCGGG CGAACTCACT GTTCGCCTTC
151 TCGCTGAGCG TGATGGCGGC GCTCACCTTC GGCTGCTTCA TCACCACCGC
201 CTTCAAAGAC AGGAGCGTCC CGGTGCGGCT GCACGTCTCG CGGATCATGC
251 TAAAAAATGT AGAAGATTTC ACTGGACCTA GAGAAAGAAG TGATCTGGGA
301 TTTATCACAT CTGATATAAC TGCTGATCTA GAGAAATATAT TTGATTGGAA
351 TGTTAAGCAG TTGTTTCTTT ATTTATCAGC AGAATATTCA ACAAAAAATA
401 ATGCTCTGAA CCAAGTTGTC CTATGGGACA AGATTGTTTT GAGAGGTGAT
451 AATCCGAAGC TGCTGCTGAA AGATATGAAA ACAAATATT TTTTCTTTGA
501 CGATGGAAT GGTCTCAAGG GAAACAGGAA TGTCACTTTG ACCCTGTCTT
551 GGAACGTCGT ACCAATGCT GGAATTCTAC CTCTTGAC AGGATCAGGA
601 CACGTATCTG TCCCATTCC AGATACATAT GAAATAACGA AGAGTTATTA
651 AATTATCTG AATTGAAAC AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

89034208:

cDNA-derived primary structure of the glycoprotein component of canine microsomal signal peptidase complex.

Peptide information for frame 1

ORF from 109 bp to 648 bp; peptide length: 180
 Category: strong similarity to known protein
 Prosite motifs: TONB_DEPENDENT_REC_1 (1-58)
 RGD (148-151)

```

1 MNTVLSRANS LFAFSLSVMA ALTFGCFITT AFKDRSVPVR LHVSRIMLKN
51 VEDFTGPRER SDLGFIISDI TADLENIFDW NVKQLFLYLS AEYSTKNNAL
101 NQVVLWDKIV LRGDNPKLL LKDMKTRYFFF DDGNGLKGNR NVTLTLSWNV
151 VPNAGILPLV TSGHVSVPF PDYIEITKSY

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64j18, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64j18, frame 1

Report for DKFZphfbr2_64j18.1

[LENGTH] 180
[MW] 20253.39
[pI] 8.66
[HOMOL] PIR:A31788 signal peptidase (EC 3.4.99.-) (SPC 22/23) - dog 1e-100
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YLR066w]
6e-15
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YLR066w] 6e-15
[PIRKW] transmembrane protein 2e-92
[PIRKW] glycoprotein 2e-92
[PIRKW] hydrolase 2e-92
[PROSITE] RGD 1
[PROSITE] MYRISTYL 2
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TONB_DEPENDENT_REC_1 1
[PROSITE] PKC_PHOSPHO_SITE 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta
[KW] SIGNAL_PEPTIDE 32

SEQ. MNTVLSRANSLFAFSLSVMAALTFGCFITTAFKDRSVPVRLHVSRIMLKNVEDFTGPRER
PRD cccccchhhhhhhhhhhhhhhhhhhhhheccccceehhhhhhhhhhhhhccccccc

SEQ. SDLGFITSDITADLENIFDWNVKQLFLYLSAEYSTKNNALNQVVLWDKIVLRGDNPKLLL
PRD ccccchhhhhhhccccccchhhhhhhhhhhhhhhccccceeeeeeccccchhhh

SEQ. KDMKTKYFFFDGNGLGKGNRVTLTSLWNVPNAGILPLVTGSGHVSVPFPDYEITKSY
PRD hhccccceeeccccccccccccceeeccccceeeccccceeecccccccccccc

Prosites for DKFZphfbr2_64j18.1

PS00001	141->145	ASN_GLYCOSYLATION	PDOC00001
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	135->141	MYRISTYL	PDOC00008
PS00013	16->27	PROKAR_LIPOPROTEIN	PDOC00013
PS00016	112->115	RGD	PDOC00016
PS00430	1->22	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKFZphfbr2_64j18.1)

DKFZphfbr2_64k24

group: transmembrane proteins

DKFZphfbr2_64k24 encodes a novel 412 amino acid protein with weak similarity to several known proteins.

The novel protein contains 5 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to AMAC1 "testicular condensing enzyme" ;
membrane regions: 5
Summary DKFZphfbr2_64k24 encodes a novel 412 amino acid protein, with
similarity to AMAC1"; product: "testicular condensing enzyme

similarity to AMAC1 "testicular condensing enzyme"

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1958 bp

Poly A stretch at pos. 1939, polyadenylation signal at pos. 1918

```

1  GGGCCCCGCT CGATTTTCCC AGGCGAGGGC ACGCCCGCGT CAGTCGCCTC
51  CCGGGCACCT TCCTCGCCAC GACACGCAGG TAACCGGGCC CCGGGAGCCG
101  GTCGGCCGGC GCGGACTGGG ACCTTGATCC TGCTGCCCCG GCCGCCCGAC
151  AAGGGAATGA GAGCGGACCC CGAACTCCAC ACACCCGCGT TTAGCCGCCA
201  CACCTAAGGG GCAGAACAGT CTTTTGGGT AAGGGCCGGG CTGGGGGCGA
251  CGGCCCCCGC CCGCTTTGCA GACTTCGGGG TGCTCTGCAC GACGCCCTGAA
301  AGGCCCGCGG GCCCGCATTT CTCTGTGCTG CCCTCCTGGA GAACCGGGAC
351  ACGGGGACGG GAGGGCCAGC ATCGGCTACG GCCCGGTTTC CCGTTTCTTT
401  CCTCTGTGCG GTCTGGGCCC TCCTGCAGCG TCCATGATGA AGGCCAGGGG
451  CTGTTGCTTT CCTCTGCCCC AGTAGCCAAC CCAAGCAAGG GAATTAATTA
501  TCTGAAGAAA TGGATACTTC TCCCTCCAGA AAATATCCAG TTAATAAACG
551  GGTGAAAATA CATCCCAACA CAGTGATGGT GAAATATACT TCTCATTATC
601  CCCAGCCTGG CGATGATGGA TATGAAGAAA TCAATGAAGG CTATGGGAAT
651  TTTATGGAGG AAAATCCAAA GAAAGGTCTG CTGAGTGAAA TGAAAAAATA
701  AGGGAGAGCT TTCTTTGGAA CCATGGATAC CCTACCTCCA CCAACAGAAG
751  ACCCAATGAT CAATGAGATT GGACAATTCC AGAGCTTTGC AGAAAAAATA
801  ATTTTTCAT CCAGAAAAT GTGGATAGTG CTGTTTGGAT CTGCTTTGGC
851  TCATGGATGT GTAGCTCTTA TCACTAGGCT TGTTCCTGAT CGGTCTAAAG
901  TTCCATCTCT AGAACTGATT TTTATCCGTT CTGTTTTTCA GGTCTTATCT
951  GTGTTAGTTG TGTGTTACTA TCAGGAGGCC CCCTTTGGAC CCAGTGGATA
1001  CAGATTACGA CTCTTCTTTT ATGGTGTATG CAATGTCATT TCTATCACTT
1051  GTGCTTATAC ATCATTTTCA ATAGTTCCTC CCAGCAATGG GACCACTATG
1101  TGGAGAGCCA CAATACAGT CTTCAGTGCC ATTTTGGCTT TTTTACTCGT
1151  AGATGAGAAA ATGGCTTATG TTGACATGGC TACAGTTGTT TGCAGCATCT
1201  TAGGTGTTTG TCTTGTCATG ATCCCAAACA TTGTTGATGA AGACAATTCT
1251  TTGTTAAATG CCTGGAAGA AGCCTTTGGG TACACCATGA CTGTGATGGC
1301  TGGACTGACC ACTGCTCTCT CAATGATAGT ATACAGATCC ATCAAGGAGA
1351  AGATCAGCAT GTGGACTGCG CTGTTTACTT TTGGTTGGAC TGGGACAAAT
1401  TGGGGAATAT CTAATATGTT TATCTTCAA GAACCCATCA TCCCATTAGA
1451  TGGAGAAACC TGGAGTTATC TCATTGCTAT ATGTGCTGT TCTACTGCAG
1501  CATCTCTAGG AGTTTATAT GCCTTGGACA AATCCATCC AGCTTTGGTT
1551  AGCACAGTAC AACATTGGGA GATTGTGGTA GCTATGGTCT TGCAGCTTCT
1601  CGTGCTGCAC ATATTTCCCTA GCATCTATGA TGTTTTTGGA GGGGTAATCA
1651  TTATGATTAG TGTTTTTGTC CTTGCTGGCT ATAACTTTA CTGGAGGAAT
1701  TTAAGAAGGC AGGACTACCA GGAATACTA GACTCTCCCA TTAATGAAT
1751  AACTGATTAT TATTGCTCA TTAATGTTCA GTTATTAATA TGTATACTGC
1801  CATTTTAAATG TTTACCTATG AATGTCTTTT GTGTTATATA ACTGACAGAG
1851  TGTATATAAA TATATAATAT ATACAAATGC AGAAAATTTA TTCTAGTCTA
1901  ATATATTCAA ATACAAATAT TAAATATATG AAATACGTTA AAAAAAATA
1951  AAAAAATA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 510 bp to 1745 bp; peptide length: 412
 Category: similarity to known protein

```

1 MDTSPSRKYP VKKRVKIHPN TVMVKYTSHY PQPGDDGYEE INEGYGNFME
51 ENPKKGLLSE MKKKGRAFFG TMDTLPPPTD DPMINEIGQF QSFAEKNIFQ
101 SRKMWIVLFG SALAHGCVAL ITRLVSDRSK VPSLELIFIR SVFQVLSVLV
151 VCYIQEAPFG PSGYRLRLEF YGVCNVSIT CAYTSFSIVP PSNGTTMWRA
201 TTTVFSAILA FLLVDEKMAV VDMATVVCSI LGVCLVMIPN IVDEDNSLLN
251 AWKEAFGYTM TVMAGLTAL SMIVYRSIKE KISMWTALFT FGWTGTIWI
301 STMFILQEPI IPLDGETWSY LIAICVCSTA AFLGVVYALD KFHPALVSTV
351 QHLEIVVAMV LQLLVLIHIFP SIYDVFGGVI IMISVFVLAG YKLYWRNLRR
401 QDYQEILDSP IK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64k24, frame 3

TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds., N = 1, Score = 191, P = 1.9e-12

TREMBL:BMAJ733_6 product: "hypothetical protein"; Bacillus megaterium bgaM gene, N = 1, Score = 137, P = 1.6e-06

PIR:G71841 hypothetical protein jhp1155 - Helicobacter pylori (strain J99), N = 1, Score = 129, P = 1.3e-05

>TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds.

Length = 362

HSPs:

Score = 191 (28.7 bits), Expect = 1.9e-12, P = 1.9e-12
 Identities = 39/105 (37%), Positives = 66/105 (62%)

Query: 289 FTFGWTGTIWIISTMFILQEPIIPLDGETWSYLIAICVCSTAFLGVVYALDKFHPALVS 348
 F FG G + + +F+LQ P++P D +WS ++A+ + + +F+ V YA+ K HPALV
 Sbjct: 248 FLFLVGLMVSVPGLFVLQTPVLPQDTLSWSCVAVGLLALVSFVCVSYAVTKAHPALVC 307

Query: 349 TVQHLEIVVAMVLQLLVLH--IFPSIYDVFGGVIIMISVFVLAGYKL 393
 V H E+VVA++LQ VL+ + PS D+ G +++ S+ ++ L
 Sbjct: 308 AVLHSEVVVALMLQYYVLYETVAPS--DIMAGVVLGSIITAQN 352

Pedant information for DKFZphfbr2_64k24, frame 3

Report for DKFZphfbr2_64k24.3

```

{LENGTH}      412
{MW}           46449.87
{pI}           6.99
{HOMOL}        TREMBL:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus
musculus testicular condensing enzyme (AMAC1) mRNA, complete cds. 8e-14
{PROSITE}      MYRISTYL 6
{PROSITE}      CK2_PHOSPHO_SITE 3
{PROSITE}      PKC_PHOSPHO_SITE 4
{PROSITE}      ASN_GLYCOSYLATION 1
{KW}           TRANSMEMBRANE 5

```

SEQ MDTSPSRKYPVKKRVKIHPNTVMVYKTSHPYQPGDDGYEENEGYGNFMEENPKKGLLSE

```

PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccchhhh
MEM      .....

SEQ      MKKKGRAFFGTMOTLPPTEDPMINEIGQFQSFAEKNIFQSRKMWIVLFGSALAHGCV
PRD      hhhhccceccccccccccccccccccccccccchhhhhhhhhccceeeeeeecccccchhhh
MEM      .....

SEQ      ITRLVSDRSKVPSELEIFIRSVFQVLSVLVVCYYQEAFFGPSGYRLRLFFYGVCNVISIT
PRD      chhhhhccccccccchhhhhhhhhhhheeeeeccccccccceeeeeeeccceeeeeee
MEM      .....MMMMMMMMMMMMMMMMMMMM.....

SEQ      CAYTSFSIVPPSNGTTMRWRTTTFVSAILAFLLVDEKMAYVDMATVVCVSILGVCLVMI
PRD      eccceeeccccccccceeeehhhhhhhhhhhhhhhhhheeeeeeeeeeeeeeeeeeeccc
MEM      .....

SEQ      IVDEDNSLLNAWKEAFGYTMTVMAGLTALSMIVYSIKEKISMWALTFTFGWTGTIWI
PRD      cccccchhhhhhhhhhhheeeeeehhhhhhhcchhhhhhhhhhhccccccccceeeec
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      STMFILQEPIIPLDGETWSYLIAICVSTAALFVGYYALDKFHPALVSTVQHLEIVVAMV
PRD      ceeeeeccccccccccccccccccchhhhhhhhhhhccccccccccchhhhhhhhhhhhhhh
MEM      MMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMM

SEQ      LQLLVLHIFPSIYDVFVGVIIMISVFLVAGYKLYWRNLRRQDYQEILDSPIK
PRD      hhhhhhhhhhhccccccccceeeeeeeccccccccchhhhhhhhhhhhhhhhhhhcccc
MEM      MMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

```

Prosites for DKFZphfbr2_64k24.3

PS000001	193->197	ASN_GLYCOSYLATION	PDOC000001
PS000005	6->9	PKC_PHOSPHO_SITE	PDOC000005
PS000005	101->104	PKC_PHOSPHO_SITE	PDOC000005
PS000005	126->129	PKC_PHOSPHO_SITE	PDOC000005
PS000005	277->280	PKC_PHOSPHO_SITE	PDOC000005
PS000006	92->96	CK2_PHOSPHO_SITE	PDOC000006
PS000006	277->281	CK2_PHOSPHO_SITE	PDOC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDOC000006
PS000008	70->76	MYRISTYL	PDOC000008
PS000008	88->94	MYRISTYL	PDOC000008
PS000008	110->116	MYRISTYL	PDOC000008
PS000008	265->271	MYRISTYL	PDOC000008
PS000008	295->301	MYRISTYL	PDOC000008
PS000008	334->340	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphfbr2_64k24.3)

DKFZphfbr2_6a17

group: brain derived

DKFZphfbr2_6a17 encodes a novel 100 amino acid protein with very weak similarity to human finger protein zFOC1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1424 bp

Poly A stretch at pos. 1405, polyadenylation signal at pos. 1389

```
1 GGGACTGAGG GGGTGGGCTT ACTCCCTGGG CAGTCTTGGG GGCCAGAGCT
51 GAGGCCAGTC CATATTACAG TGGCTGGGCT GTTTTTTCA GTAGCCCCTA
101 GCATTGGCTG GGATTCTCTG TCCTGGGTGC GCCTCCACCT CCCTTCTGAT
151 GCTTCCTGGC TATGGTGGGG TGGGAACCTC AGTTTCCCCC AAAGTCTTCC
201 CTGGATGCTG GCTTCAGGTT GAAGACCCTG GTTCTTCCAG TTCCTCACGG
251 GTTAGGTAGG GGCTCCTGCA TCACCTTCAG AATCAGTTCC AACCCCCACT
301 CTCCTTAGGC TTTGTGCTCT GCTCTGCCCT GCCAGGCTGC CCTTGTCAT
351 GTGAGTAGCA TGGGCGGGTG GTGGGGACGG CAGTGGTGAT GAAGGGGGTG
401 CACCACAGGC CTCATGAAGC AGTTCACACA TGGGCGGTG GCTGGGGCGT
451 GGCCACCACA GAGCACATGG CTGTGTCTAG GCGCAAGCAC TTTAGCAGTA
501 TCTGTTTACA TGGCAAGGA TCAAGCCGAC TACCTGTGCT GTCTACTGGG
551 ACAGCAGTCT CCGAGCTACT CCGTACCTCC CTCTGCCAGG TCGTGGAGTT
601 AGGCCCCAGT CCCTACTTGT CACTGGTTCC CACTGTGCTC CTAAGTGTGC
651 AGCACCTGGG AGCTCTGGCC TGGGGCTGGA GGCCCTGGTA GGAGCTGCAG
701 TTGGAGGCCG TTCTGTGCCC AGCAGCGGTG AGCGGCTCCC ATGGGCCCTG
751 TGTCTGCAGG GAGCCAGGGC TGGGCGACAT GTGCTGTGAA ACTGGCACCC
801 ACCTGGCGTG CTGCTGCCGC CACTTGCTTC CTGCAGCACC TCCTACCCCTG
851 CTCCTGTGTC TCCTCTCCCG CGCGCCTGGC TCAGGAGTGC TGGAAAAGCT
901 CACGCTCGG CCTGGGAGCC TGGCCTCTTG ATATACCTCG AGCTTCCCCT
951 GTGCTCCCCA GCCCCAGGAC CACTGGCCCC TTGGCCTGAG GGGCTGGGGG
1001 CCCCACGACC TGCAGCGTCG AGTCCGGGAG AGAGCCCGGA GCGGCGTGCC
1051 ATCTCGGCTC GGCCTTGCTG AGAGCCTCCG CCCTGGCTTT CTCCTGTCT
1101 GGTTCAGTGT GCTCACGTTG GTGCTACACA GCTAGAATAG ATATATTTAG
1151 AGAGAGAGAT ATTTTAAAGA CAAAGCCAC AATTAGCTGT CCTTTAACAC
1201 CGCAGAACCC CCTCCCAGAA GAAGAGCGAT CCTCGGACG GTCCGGGCGG
1251 GCACCCTCAG CCGGGCTCTT TGCAGAAGCA GCACCCTGA CTGTGGGCC
1301 GGCCCTCAGA TGTGTACATA TACGGCTATT TCCTATTTTA CTGTTCTTCA
1351 GATTTAGTAC TTGTAAATAA ACACACACAT TAAGGAGAGA TTAAACATT
1401 TTGCCAAAAA AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 389 bp to 688 bp; peptide length: 100
Category: putative protein

```
1 MKGVHHRPHE AVPTWACGWG VATTEHMAVS RRRKFSSICL HAQSSRLPV
51 LSTGTAVSEL LRTSLCQVVE LGPSPYLSLV PTVLLTVQHL GALAWGWRPW
```

BLASTP hits

Entry S70007 from database PIR:
finger protein zFOC1 - human (fragment)
Length = 183
Score = 62 (21.8 bits), Expect = 0.24, Sum P(2) = 0.22
Identities = 18/47 (38%), Positives = 24/47 (51%)

Alert BLASTP hits for DKFZphfbr2_6a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6a17, frame 2

Report for DKFZphfbr2_6a17.2

[LENGTH] 100
[MW] 10944.82
[pI] 9.49
[PROSITE] MYRISTYL 2
[PROSITE] PKC_PHOSPHO_SITE 2
[KW] Alpha_Beta

SEQ MKGVHHRPHEAVPTWACGWGVATTEHMAVSRKHFSSICLHAQGSSRLPVLSTGTAVSEL
PRD cccccccccccccccccccccchhhhhhhhhccccceccccceccccchhhh

SEQ LRTSLCQVVELGSPYLSLVPTVLLTVQHLGALAWGWRPW
PRD hhhhheeeccccceecchhhhhhhchhhhcccc

Prosite for DKFZphfbr2_6a17.2

PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	54->60	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_6a17.2)

DKFZphfbr2_6b24

group: metabolism

DKFZphfkd2_6b24 encodes a novel 334 amino acid protein with similarity to several bacterial dTDP-4-dehydroxamnose reductases (EC 1.1.1.133).

The novel protein seems to be a human enzyme similar to dTDP-4-dehydroxamnose reductases. EC 1.1.1.133 catalyses the reaction: dTDP-6-deoxy-L-mannose + NADP(+) \rightleftharpoons dTDP-4-dehydro-6-deoxy-L-mannose + NADPH.

The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

similar to dTDP-6-deoxy-L-mannose-dehydrogenases

complete cDNA, EST hits, complete cds
Nucleotide sugars metabolism seems to be a dehydrogenase
localisation: region of primer A missing

Sequenced by AGOWA

Locus: /map="5"

Insert length: 2054 bp

Poly A stretch at pos. 2028, polyadenylation signal at pos. 2015

```
1 GGGGGAGGCC CGCGTCGATC CTGGGTTGGA GGAGGTGGCG GCCGCTGAGG
51 CTGCGGCGTG AAGACGGCGG GCATGGTGGG GCGGGAGAAA GAGCTCTCTA
101 TACACTTTGT TCCCGGGAGC TGTGGGCTGG TGGAGGAGGA AGTTAACATC
151 CCTAATAGGA GGGTTCTGGT TACTGGTGCC ACTGGGCTTC TTGGCAGAGC
201 TGTACACAAA GAATTTTCAGC AGAATAATTG GCATGCAGTT GGCTGTGGTT
251 TCAGAAGAGC AAGACCAAAA TTTGAACAGG TTAATCTGTT GGATTCTAAT
301 GCAGTTCATC ACATCATTCA TGATTTTCAG CCCCATGTTA TAGTACATTG
351 TGCAGCAGAG AGAAGACCAG ATGTTGTAGA AAATCAGCCA GATGCTGCCT
401 CTCAACTTAA TGTGGATGCT TCTGGGAATT TAGCAAAGGA AGCAGCTGCT
451 GTTGGAGCAT TTCTCATCTA CATTAGCTCA GATTATGTAT TTGATGGAAC
501 AAATCCACCT TACAGAGAGG AAGACATACC AGCTCCCTTA AATTTGTATG
551 GCAAAACAAA ATTAGATGGA GAAAAGGCTG TCCTGGAGAA CAATCTAGGA
601 GCTGCTGTTT TGAGGATTCC TATTCTGTAT GGGGAAGTTG AAAAGCTCGA
651 AGAAAGTGCA GTGACTGTTA TGTTTGATAA AGTGCAGTTC AGCAACAAGT
701 CAGCAAAACAT GGATCACTGG CAGCAGAGGT TCCCCACACA TGTCAAAGAT
751 GTGGCCACTG TGTGCCGGCA GCTAGCAGAG AAGAGAAATG TGGATCCATC
801 AATTAAGGGA ACCTTTCACT GGTCTGGCAA TGAACAGATG ACTAAGTATG
851 AAATGGCATG TGCAATTGCA GATGCCTTCA ACCTCCCTAG CAGTCACTTA
901 AGACCTATTA CTGACAGCCC TGTCTTAGGA GCACAACGTC CGAGAAATGC
951 TCAGCTTGAC TGCTCCAAAT TGGAGACCTT GGGCATTGGC CAACGAACAC
1001 CATTTGCAAT TGGAAATCAA GAATCACTTT GGCCTTTCTT CATTGACAAG
1051 AGATGGAGAC AAACGGTCTT TCATTAGTTT ATTTGTGTTG GGTCTCTTTT
1101 TTTTAAAAAT GAAAAGTATA GTATGTGGCC CTTTTTAAAG AACAAAGGAA
1151 ATAGTTTTGT ATGAGTACTT TAATTGTGAC TCTTAGGATC TTTCAGGTAA
1201 ATGATGCTCT TGCAGTAGTG AAATTTGTCTA AAGAAACTAA AGGGCAGTCA
1251 TGCCCTGTTT GCAGTAATTT TTCTTTTATC CATTATGTTT GTCCTGGCTA
1301 AACTTTGGAT TTGAGTATAG TAAATTATGA TCCTTAAATA TTTGAGGGTC
1351 AGGATGAAGC AGATCTGCTG TAGACTTTTC AGATGAAATT GTTCATTCTC
1401 GTAACCTCCA TATTTTCAGG ATTTTGAAG CTGTTGACCA TTTCATGTTG
1451 ATTATTTTAA ATTGTGTGGA ATAGTATAAA AATCATTGGT GTTCATTATT
1501 TGCTTTGCCT GAGCTCAGAT CAAAATGTTT GAAGAAAGGA ACTTTATTTT
1551 TGCAAGTTAC GTACAGTTT TATGCTTGAG ATATTTCAAC ATGTTATGTA
1601 TATTGGAAC TCTACAGCTT GATGCCTCCT GCTTTTATAG CAGTTTATGG
1651 GGAGCACTTG AAAGAGCGTG TGTACATGTA TTTTCTTCT AGGCAACATC
1701 TGAATGCAAA CGTGTATTTT TTTAATATAA ATATATAACT GTCCTTTTCA
1751 TCCCATGTTG CCGCTAAGTG ATATTTTATA TGTGTGGTTA TACTCATAAT
1801 AATGGGCCTT GTAAGTCTTT TCACCATTCA TGAATAATAA TAAATATGTA
1851 CTGCTGGCAT GTAATGCTTA GTTTTCTTGT ATTTACTTCT TTTTAAAA
1901 TGTAAGGACC AAACCTCTAA ACTAATTGTT CTTTGTGTGC TTTAATTTT
1951 AAAAAATACA TTCTTCTGAT GTAACATGTG ATACATACAA AAGAATATAG
2001 TTTAATATGT ATTGAAATAA AACACAATAA AATTAAAAAA AAAAAAAA
2051 AAAA
```

BLAST Results

Entry G37115 from database EMBL:
SHGC-56899 Human Homo sapiens STS genomic.
Score = 446, P = 4.6e-14, identities = 90/91

Medline entries

99109950:
The metabolism of 6-deoxyhexoses in bacterial and animal cells.

Peptide information for frame 1

ORF from 73 bp to 1074 bp; peptide length: 334
Category: similarity to known protein

```

1  MVGREKELSI HFVPGSCRLV EEEVNIPNRR VLVGTATGLL GRAVHKEFQQ
51 NNWHA VGCGF RRARPKEFQV NLLDSNAVHH IINHDFQPHVI VHCAAERRPD
101 VVENQPDAA S QLNVDASGNL AKEAAVGAFLIYISSDYVF DGTNPPYREE
151 DIPAPLNLYG KTKLDGEKAV LENNLGA AVL RIPILYGEVE KLEESAVTVM
201 FDKVQFSNKS ANMDHWQQR F PTHVKDVATV CRQLAEKRML DPSIKGTFHW
251 SGNEQMTKYE MACAIADAFN LPSSHLPIT DSPVLGAQRP RNAQLDCSKL
301 ETLGIGQRT P FRIGIKESLW PFLIDKRWRQ TVFH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6b24, frame 1

PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans, N = 1, Score = 293, P = 6.4e-26

TREMBL:SSU51197_21 gene: "rhsD"; product:
"dTDP-6-deoxy-L-mannose-dehydrogenase"; Sphingomonas S88 sphingan
polysaccharide synthesis (spsG), (spsS), (spsR), glycosyl transferase
(spsQ), (spsI), glycosyl transferase (spsK), glycosyl transferase
(spsL), (spsJ), (spsF), (spsD), (spsC), (spsE), Urf 32, Urf 26,
ATP-binding cassette trans., N = 1, Score = 291, P = 1e-25

SWISSPROT:RFB D RHISN PROBABLE DTDP-4-DEHYDRORHAMNOSE REDUCTASE (EC
1.1.1.133) (DTDP-4-KETO- L-RHAMNOSE REDUCTASE) (DTDP-6-DEOXY-L-MANNOSE
DEHYDROGENASE) (DTDP-L- RHAMNOSE SYNTHETASE)., N = 1, Score = 283, P =
7.4e-25

>PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans
Length = 294

HSPs:

Score = 293 (44.0 bits), Expect = 6.4e-26, P = 6.4e-26
Identities = 89/276 (32%), Positives = 151/276 (54%)

```

Query: 30 RVLVTGATGLLGRAVHKEFQQNNWHA VGCGFRRARPKEFQVNNLLDSNAVHHIINHDFQPHV 89
      R+L+TGA G LGR++ K   N + V           F ++++ + + V II F+P+V
Sbjct: 3 RLLITGAGGQLGRSLAKLLVDNGRYEV-----LALDFSELDITNKMVFSSIIDSFKPNV 56

Query: 90 IVHCAAERRPDVVENQPDAA S QLNVDASGNLAKEAAVGAFLIYISSDYVFDG-TNPPYR 148
      I++ AA   D E + +A +NV   LA+ A   + +++S+DYVFDG + Y+
Sbjct: 57 IINAAAYTSVDQAELEVSSAYSVNVRGVQYLAEAAIRHNSAILHVSTDYVFDGYKSGKYK 116

Query: 149 EEDIPAPLNLYGKTKLDGEKAVLENNLGA AVL RIPILYGEVEKLEESAVTVMFDKVQFSN 208
      E DI  PL +YGK+K +GE+ +L + + +LR   +GE   + V M ++ +
Sbjct: 117 ETDI IHPLCVYGKSKAAGERLLLTLSPKSIILRTSWTFGEYGN---NFVKTML-RLAKNR 172

Query: 209 KSNAMDHWQQRFP THVKDVATVCRQLAEKRMLDPSIK-GTFHWSGNEQMTKYEMACAIAD 267
      +   Q  PT+  D+A+V  Q+A EK ++  ++K G +H++G  ++ Y+ A AI D
Sbjct: 173 DILGVVADQIGGPTYSGDIASVLIQIAEKIIVGETVKYGIYHFTGEPVSWYDFAIAIFD 232

Query: 268 AF-----NLPSSHLPITDSPVLGAQRP RNAQLDCSKLE-TLGI 305
      N+P +   D P L A+RP N+ LD +K++   GI
Sbjct: 233 EAVAQKVLNVPLVNAITADYPTL-AKRPANSCDLTKIQQA FGI 277

```


Pedant information for DKFZphfbr2 6b24, frame 1

Report for DKFZphfbr2_6b24.1

```
[LENGTH]      334
[MW]           37551.98
[pI]           6.90
[HOMOL]        PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans 6e-25
[FUNCAT]        01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YGL001c]
6e-04
[EC]            1.1.1.133 dTDP-4-dehydrorhamnose reductase 2e-16
[PIRKW]         lipopolysaccharide biosynthesis 2e-16
[PIRKW]         NADP 2e-16
[PIRKW]         oxidoreductase 2e-16
[PIRKW]         streptomycin biosynthesis 1e-19
[SUPFAM]        dTDP-dihydrostreptose synthase 1e-20
[PROSITE]       MYRISTYL 1
[PROSITE]       CK2_PHOSPHO_SITE 4
[PROSITE]       PKC_PHOSPHO_SITE 3
[PROSITE]       ASN_GLYCOSYLATION 1
[KW]            Alpha Beta
```

SEQ MVGREKELSIHFVPGSCLVEEEVNIPNRRVLVTGATGLLGRAVHKEFQNNWHAVGCGF
PRD cccccceeecccccceeeeccccccccceeeccccchhhhhhhhhhccceeeec

SEQ RRARPKFEQVNLDSNAVHHI IHDFQPHVIVHCAAERRPDVVENQPDAAASQLNVDASGNL
PRD cccccccccccccchhhhhhhhhhhccceeehhhhhhhhhhhhhhhhhhhhhhccchhh

SEQ AKEAAAVGAFLIYISSDYVFDGNTPPYREEDIAPAPLNLYGKTKLDGEKAVLENNLGAAVL
PRD hhhhhhhhhheeeeeccccccccccccccccccccccccccccchhhhhhhhhccccceeee

SEQ RIPILYGEVEKLEESAVTMFDKVFQSNKSANMDHWQQRFPTHVKDVAIVCRQLAEKRML
PRD eeeeeeeccccccchhhhhhhhhhhhhccceeeccccccccchhhhhhhhhhhhhhhhh

SEQ DPSIKGTFHWSGNEQMTKYEMACAIADAFNLPSSHLRPITDSPVLGAQRPRNAQLDCSKL
PRD cccccceeecccccchh

```
SEQ      ETLGIGQRTPFIRIGIKESLWFLIDKRWRQTVEH  
PRD      hhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
```

Prosites for DKFZphfbr2 6b24.1

PS00001	208->212	ASN_GLYCOSYLATION	PDOC00001
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	251->255	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	298->302	CK2_PHOSPHO_SITE	PDOC00006
PS00008	314->320	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2 6b24.1)

DKFZphfbr2_6i20

group: brain derived

DKFZphfbr2_6i20 encodes a novel 296 amino acid protein with similarity to ribosomal protein L15 precursor of *S. cerevisiae* mitochondria.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ribosomal protein L15 precursor, mitochondrial

complete cDNA, complete cds, EST hits
potential mitochondrial L15 ribosomal protein

Sequenced by AGOWA

Locus: /map="377.5 cR from top of Chr8 linkage group"

Insert length: 1122 bp

Poly A stretch at pos. 1099, polyadenylation signal at pos. 1071

```

1 GGGGGCCCTT GAAAGTTCTT GGATCTGCGG GTTATGGCCG GTCCCTTGCA
51 GGGCGGTGGG GCCCGGGCCC TGGACCTACT CCGGGGCTG CCGCGTGTGA
101 GCCTGGCCAA CTTAAGCCG AATCCCGGCT CCAAGAAACC GGAGAGAAGA
151 CCAAGAGGTC GGAGAAGAGG TAGAAAATGT GGCAGAGGCC ATAAAGGAGA
201 AAGGCCAAGA GGAACCCGGC CCGCTTGGG CTTTGAGGGA GGCCAGACTC
251 CATTTTACAT CCGAATCCCA AAATACGGGT TTAACGAAGG ACATAGTTTC
301 AGACGCCAGT ATAAGCCTAT GAGTCTCAAT AGACTGCAGT ATCTTATTGA
351 TTTGGGTCGT GTTGATCCTA GTCAACCTAT TGACTTAACC CAGCTTGTCA
401 ATGGGAGAGG TGTGACCATC CAGCCACTTA AAAGGGATTA TGATGTCCAG
451 CTGGTTGAGG AGGGTGCTGA CACCTTTACG GCAAAAGTTA ATATTGAAAGT
501 ACAGTTGGCT TCAGAACTAG CTATTGCTGC CATTGAAAAA AATGGTGGTG
551 TTGTTACTAC AGCCTTCTAT GATCCAAGAA GTCTGGACAT TGTATGCAAA
601 CCTGTTCAT TCTTTCTTCG TGGACAACCC ATTCCAAAAA GAATGCTTCC
651 ACCAGAAGAA CTGGTACCAT ATTACACTGA TGCAAAGAAC CGTGGGTACC
701 TGGCGGATCC TGCCAAATTT CCTGAAGCAC GACTTGAACT CGCCAGGAAG
751 TATGGTTATA TCTTACCTGA TATCACTAAA GATGAACTCT TCAAAATGCT
801 CTGTACTAGG AAGGATCCAA GGCAGATTTT CTTTGGTCTT GCTCCAGGAT
851 GGGTGGTGAA TATGGCCGAT AAGAAAATCC TAAAACCTAC AGATGAAAAT
901 CTCCTTAAGT ATTATACCTC ATGAATTCCC GTCCAAGGAA GCAGAGTTGT
951 TAAAGAGTAC TGAATAGGG GCTGAAGGAT CTATATTCCC TTATTGCATT
1001 TTCCTTATGT ATAATTTTCC AGATGGTGAT GTTACTTTTC AGTGTACTCA
1051 TATGTCTCAT TTTCACTTAA AATTAAATGG CAGGAAACAA GGAATGCATA
1101 GAGAAAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HS500354 from database EMBL:
human STS WI-12392.
Length = 426
Minus Strand HSPs:
Score = 1791 (268.7 bits), Expect = 1.1e-74, P = 1.1e-74
Identities = 375/384 (97%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 34 bp to 921 bp; peptide length: 296
Category: strong similarity to known protein

1 MAGPLQGGGA RALDLLRGLP RVSLANLKPN PGSKKPERRP RGRRRGRKCG

51 RGHKGERQRG TRPRLGFEGG QTPFYIRIPK YGFNEGHSFR RQYKPSLNR
 101 LQYLIDLGRV DPSQPIDLTQ LVNDRGVTIQ PLKRDYDVQL VEEGADTFTA
 151 KVNIEVQLAS ELAIAAIEKN GGVVTTAFYD PRSLDIVCKP VPFFLRGQPI
 201 PKRMLPPEEL VPYYTDAKNR GYLADPAKFP EARLELARKY GYLDPDITKD
 251 ELFKMLCTRK DPRQIFFGLA PGWVVNMADK KILKPTDENL LKYYTS

BLASTP hits

Entry S63258 from database PIR:
 ribosomal protein L15 precursor, mitochondrial - yeast (*Saccharomyces cerevisiae*)
 Length = 322
 Score = 259 (91.2 bits), Expect = 2.0e-22, P = 2.0e-22
 Identities = 71/200 (35%), Positives = 106/200 (53%)

Entry H70161 from database PIR:
 ribosomal protein L15 (rplO) - Lyme disease spirochete
 Length = 145
 Score = 173 (60.9 bits), Expect = 4.8e-13, P = 4.8e-13
 Identities = 45/140 (32%), Positives = 73/140 (52%)

Alert BLASTP hits for DKFZphfbr2_6i20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6i20, frame 1

Report for DKFZphfbr2_6i20.1

[LENGTH] 296
 [MW] 33495.98
 [pI] 9.98
 [HOMOL] TREMBL:AF067212_1 gene: "F37F2.1"; *Caenorhabditis elegans* cosmid F37F2. 1e-38

[FUNCAT] 05.01 ribosomal proteins [S. cerevisiae, YNL284c] 7e-15
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YNL284c] 7e-15
 [FUNCAT] j mrna translation and ribosome biogenesis [M. genitalium, MG169] 1e-06
 [BLOCKS] BL00475D
 [BLOCKS] BL00475B Ribosomal protein L15 proteins
 [PIRKW] ribosome 2e-13
 [PIRKW] mitochondrion 2e-13
 [PIRKW] protein biosynthesis 2e-13
 [SUPFAM] Escherichia coli ribosomal protein L15 4e-06
 [PROSITE] MYRISTYL 3
 [PROSITE] AMIDATION 2
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 4
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 12.50 %

SEQ MAGPLQGGGARALDLLRGLPRVSLANLKPNPGSKKPERRRPRGRRGRKCGRGHKGERQRG
 SEGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...
 PRD ccc

SEQ TRPRLGFEGGQTPFYIRIPKYGFNEGHSFRQYKPSLNLQYLIDLGRVDPSPIDLTQ
 SEG
 PRD ccc

SEQ LVNDRGVTIQPLKRDYDVQLVEEGADTFTAKVNIEVQLASELAIAAIEKNGGVVTTAFYD
 SEG
 PRD ecc

SEQ PRSLDIVCKPVPFFLRGQPIPKRMLPPEELVPYYTDAKNRGLADPAKFPPEARLELARKY
 SEG
 PRD ccc

SEQ GYLDPDITKDELFKMLCTRKDPRQIFFGLAPGWVVNMADKKILKPTDENLLKYYTS
 SEG
 PRD ccc

Prosites for DKFZphfbr2_6i20.1

PS00005 33->36 PKC_PHOSPHO_SITE PDOC00005
 PS00005 88->91 PKC_PHOSPHO_SITE PDOC00005

PS00005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00006	248->252	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	171->177	MYRISTYL	PDOC00008
PS00008	268->274	MYRISTYL	PDOC00008
PS00009	41->45	AMIDATION	PDOC00009
PS00009	45->49	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_6i20.1)

DKFZphfbr2_6017

group: nucleic acid management

DKFZphfbr2_6017 encodes a novel 455 amino acid protein with strong similarity to DEAD-box ATP-dependent RNA helicases YHR065c and T26G10.1.

The *S. cerevisiae* protein YHR065c is required for maturation of the 35S RNA primary transcript.

The new protein can find application in modulating rRNA maturation.

strong similar to RNA helicases

complete cDNA, complete cds, EST hits
probable start at Bp 27 matches kozak consensus ANNatgG
involved in maturation of r-RNA ??
YHR065c/Rrp3p is involved in maturation of the 35S primary transcript
Drslp cold-sensitive mutation has slow 27S to 25S pre-rRNA
conversion and is deficient in 60S ribosomal subunits

Sequenced by AGOWA

Locus: unknown

Insert length: 1840 bp

Poly A stretch at pos. 1815, polyadenylation signal at pos. 1793

```
1  GGGGACTTCC GGAGACCTCA CACAAGATGG CGGCACCCGA GGAACACGAT
51  TCTCCGACCG AAGCGTCCCA GCCGATTGTG GAAGAGGAGG AACTAAAC
101 ATTTAAAGAC CTGGGTGTGA CAGATGTGTT GTGTGAAGCT TGTGACCACT
151 TGGGATGGAC AAAACCCACC AAGATTCAGA TTGAAGCTAT TCCTTTGGCC
201 TTACAAGGTC GTGATATCAT TGGGCTTGCA GAAACTGGCT CTGGAAGAC
251 AGGCGCCTTT GCTTTGCCCA TTCTAAACGC ACTGCTGGAG ACCCCGCAGC
301 GTTTGTGTTG CCTAGTTCTT ACCCGGACTC GGGAGCTGGC CTTTCAGATC
351 TCAGAGCAGT TTGAAGCCCT GGGGTCTCTT ATTGGAGTGC AGAGTGCTGT
401 GATTGTAGGT GGAATTGATT CAATGTCTCA ATCTTTGGCC CTTGCAAAAA
451 AACCACATAT AATAATAGCA ACTCCTGGTC GACTGATTGA CCACTTGGAA
501 AATACGAAAG GTTTCAACTT GAGAGCTCTC AAATACTTGG TCATGGATGA
551 AGCCGACCGA ATACTGAATA TGGATTTTGA GACAGAGGTT GACAAGATCC
601 TCAAAGTGAT TCCTCGAGAT CGGAAAACAT TCCTCTTCTC TGCCACCATG
651 ACCAAGAAGG TTCAAAACTT TCAGCGAGCA GCTCTGAAGA ATCCTGTGAA
701 ATGTGCCGTT TCCTCTAAAT ACCAGACAGT TGAATAATTA CAGCAATATT
751 ATATTTTTAT TCCTCTAAA TTCAAGGATA CCTACCTGGT TTATATTCTA
801 AATGAATTGG CTGGAACTC CTTTATGATA TTCTGCAGCA CCTGTAATAA
851 TACCCAGAGA ACAGCTTTGC TACTGCGAAA TCTTGGCTTC ACTGCCATCC
901 CCCTCCATGG ACAATGAGT CAGAGTAAGC GCCTAGGATC CCTTAATAAG
951 TTTAAGGCCA AGGCCGTTT CATTCTTCTA GCAACTGACG TTGCCAGCCG
1001 AGGTTTGGAC ATACCTCATG TAGATGTGGT TGTCAACTTT GACATTCCCTA
1051 CCCATCCCAA GGATTACATC CATCGAGTAG GTCGAACAGC TAGAGCTGGG
1101 CGCTCCGGAA AGGCTATTAC TTTTGTGACA CAGTATGATG TGAAGTCTTT
1151 CCAGCGCATA GAACACTTAA TTGGGAAGAA ACTACCAGGT TTTCCAAAC
1201 AGGATGATGA GGTATGATG CTGACAGAAC GCCTCGCTGA AGCCCAAGG
1251 TTTGCCCGAA TGGAGTTAAG GGAGCATGGA GAAAAGAAGA AACGCTCGCG
1301 AGAGGATGCT GGAGATAATG ATGACACAGA GGGTGCTATT GGTGTCAGGA
1351 ACAAGGTGGC TGGAGGAAAA ATGAAGAAGC GGAAAGGCCG TTAATCACTT
1401 TTATGAAGGC TCGAGTTCTG CTGTTCTGTA AAAGAAAATT GGAGAATGAA
1451 ACCTGCTCCA ACAGAGATCA TGAGACTGAA ATTGGTCAGA ATTGTGTCCA
1501 GAATGTGCTC AGCTAATTCA GTATTCTTCC CCATTCTGGG TTGGAGTTTA
1551 CTGCAGAGTA ATTCTTACAG TGCTGATGTC AAGACTGTTA CTGTTCTTCG
1601 ACTTTGATTC CTTGCTCATG ACATGAGTAG GGTGTGCTCT TCTGTCACTT
1651 CACACAGACC TTTTGCCCTT TTTAGCTGCA AGTCAAGGAC TAGGTTGATG
1701 ATGCCCATGA CCTGTAATTG TAAAGAAGCT TGGACATCTG CAAATGATAT
1751 TTAAACCATC TTGGCTTGTG CTTTATTCAA ACTAATGTGA AACAATAAAT
1801 TTAATATTA TTTTAAAAG AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 1391 bp; peptide length: 455
 Category: strong similarity to known protein

```

1  MAAPEEHDSP TEASQPIVEE EETKTFKDLG VTDVLC EACD QLGWTKPTKI
51 QIEAIPALQ GRDIIGLAET GSGKTGAFAL PILNALLETQ QRLFALVLTQ
101 TRELAFQISE QFEALGSSIG VQSAVIVGGI DSMSQSLALA KKPHEIIATP
151 GRLIDHLENT KGFNLRLKY LVMDEADRIL NMDFETEVDK ILKVIPDRK
201 TFLFSATMTK KVQKLQRAAL KNPVKCAVSS KYQTVEKLQQ YYIFIPSKFK
251 DTYLVYILNE LAGNSFMIFC STCNNTQRTA LLLRNLGFTA IPLHGQMSQS
301 KRLGSLNKKF AKARSILLAT DVASRGDIP HVDVVVNFDI PTHSKDYIHR
351 VGR TARAGRS GKAITFVTQY DVELFQRIEH LIGKKLPGFP TQDDEVMMLT
401 ERVAEAQRFA RMELREHGEK KKRSREDAGD NDDTEGAIGV RNKVAGGKMK
451 KRKGR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6ol7, frame 3

PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - *Caenorhabditis elegans*, N = 1, Score = 1497, P = 1.6e-153

PIR:S46713 hypothetical protein YHR065c - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 1154, P = 3.6e-117

TREMBL:ATH010462_1 gene: "RH10"; product: "RNA helicase"; *Arabidopsis thaliana* mRNA for DEAD box RNA helicase, RH10, N = 1, Score = 1122, P = 8.9e-114

TREMBL:AC002985_2 product: "R27090_2"; Human DNA from chromosome 19-specific cosmid R27090, genomic sequence, complete sequence., N = 1, Score = 950, P = 1.5e-95

>PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - *Caenorhabditis elegans*

Length = 489

HSPs:

Score = 1497 (224.6 bits), Expect = 1.6e-153, P = 1.6e-153
 Identities = 283/442 (64%), Positives = 364/442 (82%)

```

Query:   19 EEEETKTFKDLGVTDVLC EACDQLGWTKPTKIQIEAIPALQGRDIIGLAETGSGKTGAF 78
          E+ + K+F +LGV+ LC+AC +LGW KP+KIQ A+P ALQG+D+IGLAETGSGKTGAF
Sbjct:   39 EDVKEKSFAELGVSQPLCDACQRLGWMKPSKIQQAALPHALQGDVIGLAETGSGKTGAF 98

Query:   79 ALPILNALLETQRLFALVLTPTRELAFQISEQFEALGSSIGVQSAVIVGGIDSMSQSLA 138
          A+P+L +LL+ PQ F LVLTPTRELAFQI +QFEALGS IG+ +AVIVGG+D +Q++A
Sbjct:   99 AIPVLQSLLDHPQAFFCLVLTPTRELAFQIGQQFEALGSGIGLIAAVIVGGVDMAAQAMA 158

Query:   139 LAKKPHEIIATPGRLIDHLENTKGFNLRLKYLVMDEADRILNMDFETEVDKILKVIPRD 198
          LA++PHII+ATPGRL+DHELENTKGFNL+ALK+L+MDEADRILNMDFE E+DKILKVIPR+
Sbjct:   159 LARRPHIIVATPGRLVDHLENTKGFNLKALKFLIMDEADRILNMDFEVELDKILKVIPRE 218

Query:   199 RKTFLSATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQQYYIFIPSKFKDTYLVYIL 258
          R+T+LFSATMTKKV KL+RA+L++P + +VSS+Y+TV+ L+Q+YIF+P+K+K+TYLVY+L
Sbjct:   219 RRTYLFSAATMTKKVSKLERASLRDPARVSVSSRYKTVDNLKQHYIFVPNKYKETYLVYLL 278

Query:   259 NELAGNSFMIFCSTCNNTQRTALLLRNLGFTAIPLHGQMSQSKRLGSLNKKFAKARSILL 318
          NE AGNS ++FC+TC T + A++LR LG A+PLHGQMSQ KRLGSLNKKF+KAR IL+
Sbjct:   279 NEHAGNSAIVFCATCATMQIAVMLRQLGMQAVPLHGQMSQEKRLGSLNKKFKSKAREILV 338

Query:   319 ATDVASRGDIPHDVVVNFDIPTHSKDYIHRVGR TARAGRS GKAITFVTQYDVLEFQRI 378
          TDVA+RGDIPHDV+V+N+D+P+ SKDY+HRVGR TARAGRS GAIT VTQYDVE +Q+I
Sbjct:   339 CTDVAARGLDIPHDVMVINYDMPQS KQDYVHRVGR TARAGRS GAITVVTQYDVEAYQKI 398

Query:   379 EHLIGKKLPGFPPTQDDEVMMLTERVAEAQRFA RMELREHGEKKK-----RSREDAGDND 433
          E +GKKL + ++EVM+L ER EA AR+E++E EKKK R +D GD ++
Sbjct:   399 EANLGKKLDEYKCVENEVMVVERTQEATENARIEMKEMDEKKKSGKKRRQNDDFGDTTEE 458

Query:   434 TEGAIGVRNKVAGGKMKKRKGR 455

```

+ G + K GG+ GR
 Sbjct: 459 SGGRFKMGIKSMGGRGGSGGGR 480

Pedant information for DKFZphfbr2_6ol7, frame 3

 Report for DKFZphfbr2_6ol7.3

[LENGTH] 455
 [MW] 50646.80
 [pI] 9.18
 [HOMOL] PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - Caenorhabditis elegans
 1e-167
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YHR065c] 1e-127
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YHR065c] 1e-127
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YHR169w] 2e-79
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 1e-71
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 4e-66
 [FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 1e-63
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 1e-58
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YDL084w] 1e-55
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YOR204w] 5e-55
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 5e-55
 [FUNCAT] l genome replication, transcription, recombination and repair [H.
 influenzae, HI0892] 9e-48
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR276c] 2e-45
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 4e-42
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 7e-16
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YMR190c] 7e-12
 [FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 7e-12
 [FUNCAT] r general function prediction [M. jannaschii, MJ1401] 5e-06
 [BLOCKS] BL00175B Phosphoglycerate mutase family phosphohistidine proteins
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-60
 [PIRKW] RNA binding 7e-69
 [PIRKW] DEAD box 7e-69
 [PIRKW] transmembrane protein 9e-41
 [PIRKW] DNA binding 3e-55
 [PIRKW] recF recombination pathway 3e-11
 [PIRKW] ATP 1e-126
 [PIRKW] purine nucleotide binding 7e-69
 [PIRKW] P-loop 1e-126
 [PIRKW] hydrolase 1e-55
 [PIRKW] protein biosynthesis 7e-69
 [PIRKW] ATP binding 3e-61
 [SUPFAM] ATP-dependent RNA helicase eIF-4A 8e-06
 [SUPFAM] WW repeat homology 4e-58
 [SUPFAM] translation initiation factor eIF-4A 7e-69
 [SUPFAM] DEAD/H box helicase homology 1e-126
 [SUPFAM] recQ helicase homology 5e-12
 [SUPFAM] ATP-dependent RNA helicase homology 8e-06
 [SUPFAM] unassigned DEAD/H box helicases 1e-126
 [SUPFAM] ATP-dependent RNA helicase DBP1 4e-60
 [SUPFAM] ATP-dependent RNA helicase DHH1 1e-58
 [SUPFAM] recQ protein 3e-11
 [SUPFAM] tobacco ATP-dependent RNA helicase DB10 4e-58
 [SUPFAM] Bloom's syndrome helicase 5e-12
 [PROSITE] DEAD ATP HELICASE 1
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 5
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 6
 [PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Helicases conserved C-terminal domain
 [PFAM] DEAD and DEAH box helicases
 [KW] Alpha_Beta

SEQ MAAPEEHDSPTASQPIVEEEETKTFKDLGVTDLVCEACDQLGWTKPTKIQIEAIPALQ
 PRD cccccccccccccchhhhhhhhhhhccccchhhhhhhhhhhcccccccccccccccccc

SEQ GRDIIGLAETSGKGTAFALPILNALLETPQRLFALVLTPTRELAFQISEQFEALGSSIG
 PRD cceeeccccccccceehhhhhhhhhccccceeeccccchhhhhhhhhhhhhhhhhhhhhcc

```

SEQ  VQSAVIVGGIDSMSSQLALAKKPHIIATPGRLIDHLENTKGFNLRAKYLVMDEADRIL
PRD  eeeeeeeccchhhhhhhhhccceeeeeeccccccccccccccccccccceehhhhhhhh

SEQ  NMDFETEVDKILKVIPRDRKTLFSAATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQO
PRD  hhcchhhhhhhhhccchhhhhhhhhccchhhhhhhhhhhccceeeeeeccccchhhhh

SEQ  YYIFIPSKFKDYLVIYILNELAGNSFMIFCSTCNNTQRTALLRLNGFTAIPLHGQMSQS
PRD  hhhhhhhhhhhhhhhhhhhhhccceeeeeeccchhhhhhhhhhhhhccceeeccccchhh

SEQ  KRLGSLNKFKAARISILLATDVASRGLDIPHDVVVNFIDIPTHSKDYIHRVGRRTARAGRS
PRD  hhhhhhhhhhhhhhhcchhhhhhhhhccccceeeeeeccccccccceeecccccccccc

SEQ  GKAITFVTQYDVELFQRIEHLIGKKLPGFPTQDDEVMLTERVAEAQRFAARMELREHGEK
PRD  cceeeeeeccchhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KKRSREDAGDNDDEGAIGVRNKVAGGKMKRKRGR
PRD  hhhhhcccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_6ol7.3

PS00001	274->278	ASN_GLYCOSYLATION	PDOC00001
PS00004	421->425	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	72->75	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	229->232	PKC_PHOSPHO_SITE	PDOC00005
PS00005	276->279	PKC_PHOSPHO_SITE	PDOC00005
PS00005	300->303	PKC_PHOSPHO_SITE	PDOC00005
PS00005	354->357	PKC_PHOSPHO_SITE	PDOC00005
PS00005	360->363	PKC_PHOSPHO_SITE	PDOC00005
PS00005	400->403	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	25->29	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	368->372	CK2_PHOSPHO_SITE	PDOC00006
PS00006	391->395	CK2_PHOSPHO_SITE	PDOC00006
PS00006	424->428	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	116->122	MYRISTYL	PDOC00008
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	128->134	MYRISTYL	PDOC00008
PS00009	382->386	AMIDATION	PDOC00009
PS00017	68->76	ATP_GTP_A	PDOC00017
PS00039	172->181	DEAD_ATP_HELICASE	PDOC00039

Pfam for DKFZphfbr2_6ol7.3

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLPpWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACAQTGSGKTAAFG ++ +++++++G++KPT+IQ +AIP++L+GRD+++ A TGSGKT+AF		
Query	30	GVTDVLCACDQLGWTKPTKIQIEAIPALQGRDIIIGLAETGSGKTGAF	78
HMM	lIPMLQHIDwdPWpqpPQdPrALILAPTRELAMQIEEcRkFgkHMngIR ++P+L ++++P + ++AL+L+PTRELA QI+E++++G++++ ++		
Query	79	ALPILNALLETp----QR-LFALVLTPTRELAFQISEQFEALGSSIG-VQ	122
HMM	ImciYGGtnMRdQMRmLeRGpPHIVIAATPGRLIDHIER.gtlDLDrIeML +++I+GG + + Q L+++P HI+IATPGRLIDH+E+ ++L++++L		
Query	123	SAVIVGGIDSMSSQLALAKKP-HIIATPGRLIDHLENTKGFNLRAKYL	171
HMM	VMDEADRMLDMGFIDQIRrIMrqIPmpwNRQTMMSATMPdeIqELARrF VMDEADR+L+M+F+ ++++I++ IP ++R T +FSATM++++Q+L+R+		
Query	172	VMDEADRILNMDFETEVDKILKVIP--RDRKTLFSAATMTKKVQKLQRAA	219
HMM	MRNPIRInIdMdElTtnEnIkQwYiyVerEMWkfcdLcrLIe* ++NP+ ++ +++++T++ ++Q+YI+++ + K +L++++		
Query	220	LKNPVKCAVSSKYQVE-KLQYYIFIP-SKFKDYLVIYILN	259

HMM_NAME Helicases conserved C-terminal domain

HMM *EileeWLknlGirvmYIHGdMpQeERdeIMddFNnGEynVLICtdVvgR


```
Query      277  ++ + L+NLG++++ +HG+M+Q +R+  +++F++  +L++TDV++R
                QRTALLLRNLGFTAIPLHGQMSQSKRLGSLNKFKAKARSILLATDVASR  325
HMM
                GIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG*
                G+DIP V++V+N+D+P ++ +YI+R+GRT+R+G
Query      326  GLDIPHDVVVNFDIPTHSKDYIHRVGRTARAG  358
```

DKFZphfbr2_71o20

group: brain derived

DKFZphfbr2_71o20 encodes a novel 232 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits
on genomic level encoded by AC006186 (3 exons)

Sequenced by GBF

Locus: /map="10q22.1"

Insert length: 1768 bp

Poly A stretch at pos. 1742, polyadenylation signal at pos. 1726

```
1 GGGGGCAGCA GGCCAAGGGG GAGGTGCGAG CGTGGACCTG GGACGGGTCT
51 GGGCGGCTCT CGGTGGTTGG CACGGGTTCC CACACCCATT CAAGCGGCAG
101 GACGCATTG TCTTAGCAGT TCTCGCTGAC CGCGCTAGCT GCGGCTTCTA
151 CGCTCCGGCA CTCTGAGTTC ATCAGCAAAC GCCCTGGCGT CTGTCTTCAC
201 CATGCCTAGC CTTTGGGACC GCTTCTCGTC GTCGTCCACC TCCTCTTCGC
251 CCTCGTCTT GCCCCGAAC CCCACCCAG ATCGGCCGCC GCGCTCAGCC
301 TGGGGGTCGG CGACCCGGGA GGAGGGGTTT GACCGCTCCA CGAGCCTGGA
351 GAGCTCGGAC TGCAGTCCC TGGACAGCAG CAACAGTGGC TTCGGGCGCG
401 AGGAAGACAC GGCTTACCTG GATGGGGTGT CGTTGCCCGA CTTCGAGCTG
451 CTCAGTGACC CTGAGGATGA AACTTGTGT GCCAACCTGA TGCAGCTGCT
501 GCAGGAGAGC CTGGCCAGG CGCGGCTGGG CTCTCGACGC CTTGCGCGCC
551 TGCTGATGCC TAGCCAGTTG GTAAGCCAGG TGGGCAAAGA ACTACTGCGC
601 CTGGCCTACA GCGAGCCGTG CGGCCTGCGG GGGGCGCTGC TGGACGTCTG
651 CGTGGAGCAG GGCAAGAGCT GCCACAGCGT GGGCCAGCTG GCACTCGACC
701 CCAGCCTGGT GCCCACCTTC CAGCTGACCC TCGTGCTGCG CCTGGACTCA
751 CGACTCTGGC CCAAGATCCA GGGGCTGTTT AGCTCCGCCA ACTCTCCCTT
801 CCTCCCTGGC TTCAGCCAGT CCCTGACGCT GAGCACTGGC TTCCGAGTCA
851 TCAAGAAGAA GCTGTACAGC TCGGAACAGC TGCCCATGGA GGAGTGTGTA
901 ACTTCAACCT GAGGGGGCCG ACAGTGCCCT CCAAGACAGA GACGACTGAA
951 CTTTGGGGT GGAGACTAGA GGCAGGAGCT GAGGGACTGA TTCCAGTGGT
1001 TGGAAAAGT AGGCAGCCAC CTAAAGTGGG GGTGGGGGAA TAGTGTTCCT
1051 CAGGAAGCTC ATTGAGTTGT GTGCGGGTGG CTGTGCATTG GGGACACATA
1101 CCCCTCAGTA CTGTAGCATG AAACAAAGGC TTAGGGGCCA ACAAGGCTTC
1151 CAGCTGGATG TGTGTGTAGC ATGTACCTTA TTATTTTGT TACTGACAGT
1201 TAACAGTGGT GTGACATCCA GAGAGCAGCT GGGCTGCTCC CGCCCCAGCC
1251 TGGCCCAGGG TGAAGGAAGA GGCACGTGCT CCTCAGAGCA GCCGGAGGGA
1301 AGGGGGAGGT CGGAGGTCGT GGAGGTGGTT TGTGTATCTT ACTGGTCTGA
1351 AGGGACCAAG TGTGTTTGT GTTTGTTTG TATCTTGTGTT TTCTGATCGG
1401 AGCATCATA CTGACCTGTT GTAGGCAGCT ATCTTACAGA CGCATGAATG
1451 TAAGAGTAGG AAGGGGTGGG TGTCAGGGAT CACTTGGGAT CTTTGACACT
1501 TGAAAAATTA CACCTGGCAG CTGCGTTTAA GCCTTCCCCC ATCGTGTACT
1551 GCAGAGTTGA GCTGGCAGGG GAGGGGCTGA GAGGGTGGGG GCTGGAACCC
1601 CTTCCCGGGA GGAGTGCCAT CTGGGTCTTC CATCTAGAAC TGTTTACATG
1651 AAGATAAGAT ACTCACTGTT CATGAATACA CTTGATGTTT AAGTATTAA
1701 ACCTATGCAA TATTTTTCAC TTTTCTAATA AACATGTTTG TTAAACAAA
1751 AAAAAAAAAA AAAAAAAA
```

BLAST Results

Entry AC006186 from database EMBLNEW:

*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 10 clone
CRI-JC2048 map 10q22.1; HTGS phase 1, 4 unordered pieces.
Score = 6512, P = 0.0e+00, identities = 1326/1345
3 exons

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 897 bp; peptide length: 232
Category: putative protein

```
1 MPSSLWDRFSS SSTSSSPSSL PRTPTPDRPP RSAWGSATRE EGFDRSTSLE
51 SSDCESLDSS NSGFGPEEDT AYLDGVSLPD FELLSDPEDE HLCANLMQLL
101 QESLAQARLG SRRPARLLMP SQLVSVQVKE LLRLAYSEPC GLRGALLDVC
151 VEQGSCHSV GQLALDPSLV PTFQLTLVLR LDSRLWPKIQ GLFSSANSFP
201 LPGFSQSLTL STGFRVIKKK LYSSEQLPIE EC
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_71o20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_71o20, frame 1

Report for DKFZphfbr2_71o20.1

```
[LENGTH]      232
[MW]           25354.60
[pI]           4.87
[PROSITE]      MYRISTYL      2
[PROSITE]      CK2_PHOSPHO_SITE      6
[PROSITE]      GLYCOSAMINOGLYCAN      1
[PROSITE]      PKC_PHOSPHO_SITE      1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      17.67 %

SEQ  MPSSLWDRFSSSSTSSSPSSLPRTPTPDRPPRSAWGSATREEGFDRSTSLESSDCESLDSS
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  NSGFGPEEDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQESLAQARLGSRRPARLLMP
SEG  XX.....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SQLVSVQVKELLRLAYSEPCGLRGALLDVCVEQGSCHSVGQLALDPSLVPTFQLTLVLR
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LDSRLWPKIQGLFSSANSFPPLPGFSQSLTLSTGFRVIKKKLYSSEQLPIEEC
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphfbr2_71o20.1

PS00002	62->66	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	3->7	CK2_PHOSPHO_SITE	PDOC00006
PS00006	38->42	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS00008	141->147	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_71o20.1)

DKF2phfbr2_72b18

group: nucleic acid management

DKF2phfbr2_72b18 encodes a novel 715 amino acid protein with similarity to E. coli DNA-damage-inducible protein dinP and other proteins induced by DNA-damage.

The novel protein is similar to dinP of E. coli, yqjH of B. subtilis, dinP of M. tuberculosis and T19K24.15 of A. thaliana. The dinB/P pathway is a second SOS-pathway in E. coli. Therefore the new gene seems to be involved in DNA repair.

The new protein can find application in modulating DNA repair and mutagenesis.

similarity to DNA damage induced genes

complete cDNA, complete cds, potential start at Bp 49, EST hits
localisation primer site B is missing!

Sequenced by LMU

Locus: /map="416.0 cR from top of Chr18 linkage group"??

Insert length: 2475 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2431

```
1 GGGGGAGGAA GCGGGCGGCG ACGACGAGGA AGACGCCGAG GCCTGGGCCA
 1 TGGAACTGGC GGACGTGGGG GCGGCAGCCA GCTCGCAGGG AGTTCATGAT
101 CAAGTGTGTC CCACACCAAA TGCTTCATCC AGAGTCATAG TACATGTGGA
151 TCTGGATTGC TTTTATGCAC AAGTAGAAAT GATCTCAAAT CCAGAGCTAA
201 AAGACAAACC TTTAGGGGTT CAACAGAAAT ATTTGGTGGT TACCTGCAAC
251 TATGAAGCTA GGAACTTGG AGTTAAGAAA CTTATGAATG TCAGAGATGC
301 AAAAGAAAAG TGCCACAGT TGGTATTAGT TAATGGAGAA GACCTGACCC
351 GCTACAGAGA AATGTCTTAT AAGGTTACAG AATTACTGGA AGAATTTAGT
401 CCACTTGTGG AGAGACTTGG ATTTGATGAA AATTTTGTGG ATCTAACAGA
451 AATGGTTGAG AAGAGACTAC AGCAGCTGCA AAGTGATGAA CTTTCTGCGG
501 TGACTGTGTC GGGTCATGTA TACAATAATC AGTCTATAAA CCTGCTTGAC
551 GTCTTGACAC TCAGACTACT TGTGGGATCT CAGATTGCAG CAGAGATGCG
601 GGAAGCCATG TATAATCAGT TGGGGCTCAC TGGCTGTGCT GGAGTGGCTT
651 CTAATAAACT GTTGGCAAAA TTAGTTTCTG GTGTCTTTAA ACCAAATCAA
701 CAAACAGTCT TATTACCTGA AAGTTGTCAA CATCTTATTC ATAGTTTGAA
751 TCACATAAAG GAAATACCTG GTATTGGCTA TAAAACTGCC AAATGTCTTG
801 AAGCACTGGG TATCAATAGT GTGCGTGATC TCCAAACCTT TTCACCCAAA
851 ATTTTAGAAA AAGAATTAGG AATTTCAAGT GCTCAGCGTA TCCAAAAGCT
901 CAGTTTGGGA GAGGATAACT CCCCTGTGAT ACTCTCAGGA CCACCTCAGT
951 CCTTTAGTGA AGAAGATTCA TTTAAAAAAT GTACATCTGA AGTTGAAGCT
1001 AAAAATAAGA TTGAAGAACT ACTTGCTAGT CTTTAAACA GAGTATGCCA
1051 AGATGGAAGG AAGCCTCATA CAGTGAGATT AATAATCCGT CGGTATTCCT
1101 CTGAGAAGCA CTATGGTCTG GAGAGTCGTC AGTGCCCTAT TCCTTCACAT
1151 GTAATTCAGA AATTAGGGAC AGGAAATTAT GATGTGATGA CCCCATGGT
1201 TGATATACTT ATGAACTTTT TTCGAAATAT GGTGAATGTG AAGATGCCAT
1251 TTCACCTTAC CCTTCTAAGT GTGTGCTTCT GCAACCTTAA AGCACTAAAT
1301 ACTGCTAAGA AAGGGCTTAT TGATTATTAT TTAATGCCAT CATTATCAAC
1351 TACTTCACGC TCTGGCAAGC ACAGTTTAA AATGAAAGAC ACTCATATGG
1401 AAGATTTTCC CAAAGACAAA GAAACAAACC GGGATTTCCCT ACCAAGTGGA
1451 AGAATTGAAA GTACAAGAAC TAGGGAGTCT CCACTAGATA CCACAAATTT
1501 TTCTAAAGAA AAAGACATTA ATGAATCCCT ACTCTGTTCA CTTCTGAAG
1551 GTGTTGACCA AGAAGTCTCC AAGCAGCTTC CAGTAGATAT TCAAGAAGAA
1601 ATCCTTTCTG GAAAATCTAG GGAATAATTT CAAGGGAAG GAAGTGTGAG
1651 TTGTCCATTA CATGCCTCTA GAGGAGTATT ATCTTTCTTT TCTAAAAAAC
1701 AAATGCAAGA TATTCCTATA AATCCTAGAG ATCATTATAT CAGTAGCAAA
1751 CAGGTATCCT CTGTATCTCC TTGTGAACCG GGAACATCAG GCTTTAATAG
1801 CAGTAGTTCT TCTTACATGT CTAGCCAAAA GGATTATTC AATTATTAG
1851 ATAATAGATT AAAAGATGAA CGAATAAGTC AAGGACCTAA AGAACCTCAA
1901 GGATTCCACT TTACAAATTC AAACCTGCT GTGTCTGCTT TTCATTCAAT
1951 TCCAAACTTG CAGAGTGAGC AACTTTTCTC CAGAAACCAC ACTACAGATA
2001 GCCATAAGCA AACAGTAGCA ACAGACTCTC ATGAAGGACT TACAGAAAT
2051 AGAGAGCCAG ATTCTGTTGA TGAGAAAATT ACTTTCCCTT CTGACATTGA
2101 TCCTCAAGTT TTCTATGAAC TACCAGAAGC AGTACAAAAG GAACTGCTGG
2151 CAGAGTGGA GAGAACAGGA TCAGATTTC ACATTGGACA TAAATAAGCA
2201 TATTCAGCAA AAAGGTCTGA AAAGCAAGGG AATACCATTA TTTTCGGATT
2251 AGCGGTTTAT TAAGCTCTTC TATATTAAAC ACTAATAGAT ATTCAATAAC
2301 GGAGTAAACT GTTCCAGATA AAGCAAGAA AGTTGCAAGA AGTAAATCT
2351 GGCACAAAGC GTAAAAATAT AACAGAAGAA ATAATGTAAA ATACTATCTT
2401 TTATGTCTAA AGCCATTTTA TATTACTTTT CAATAAAAG AATATCATGG
2451 TCAAAAAAAA AAAAAAATA AAAAC
```

BLAST Results

 Entry HS086339 from database EMBL:
 human STS WI-11064.
 Score = 1523, P = 3.0e-64, identities = 327/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2194 bp; peptide length: 715
 Category: similarity to known protein

```

1 MELADVGAAG SSQGVHDQVL PTPNASSRVI VHVLDLCFYA QVEMISNPGL
51 KDKPLGVQOK YLVVTCNYEA RKLGVKKLMN VRDAKEKCPQ LVLVNGEDLT
101 RYREMSYKVT ELLEEFSPV ERLGFDEFV DLTEMVEKRL QQLQSDLSA
151 VTVSGHVYNN QSINLLDLVH IRLLVGSQIA AEMREAMYNQ LGLTGCAGVA
201 SNKLLAKLVS GVFKPNQQT VLLPESQHLI HSLNHIKEIP GIGYKTAKCL
251 EALGINSVRD LQTFSPKILE KELGISVAQR IQKLSFGEDN SPVILSGPPQ
301 SFSEEDSFKK CTSEVEAKNK IEELLASLLN RVCQDGRKPH TVRLIIRYS
351 SEKHYGRESR QCPIPSHVI KLGTYNDVM TPMVDILMKL FRNMVNVKMP
401 FHLTLLSVCF CNLKALNTAK KGLIDYLLMP SLSTTSRSGK HSFKMKDTHM
451 EDFPKDKETN RDLPSGRIE STRTRESPLD TTNFSKEKDI NEFPLCSLPE
501 GVDQEVSKQL PVDIQEELS GKSREKFGK GSVSCPLHAS RGVLSFFSKK
551 QMQDIPINPR DHLSSSKQVS SVSPCEPGTS GFNSSSSSYM SSQKDYSYL
601 DNRLKDERIS QGPKEPQGFH FTNSNPVSA FHSFPNLQSE QLFNRHTTD
651 SHKQTVATDS HEGLTENREP DSVDEKITEP SDIDPQVFYE LPEAVQKELL
701 AEWKRTGSDF HIGHK
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72b18, frame 2

PIR:H64747 DNA-damage-inducible protein dinP - Escherichia coli, N =
 2, Score = 212, P = 4.2e-27

PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis,
 N = 2, Score = 230, P = 5.2e-26

>PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis
 Length = 414

HSPs:

Score = 230 (34.5 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
 Identities = 47/112 (41%), Positives = 73/112 (65%)

Query: 27 SRVIVHVDLCFYAQVEMISNPGLKDKPLGV-----QOKYLVVTCNYEARLGVKKLMNV 81
 SR+I H+D++ FYA VEM +P L+ KP+ V ++K +VVTC+YEAR GVK M V
 Sbjct: 5 SRIIFHIDMNSFYASVEMAYDPALRGKPVAVAGNVKERKGI VVTCSEARARGVKTMPV 64

Query: 82 RDAKEKCPQLVLVNGEDLTRYREMSYKVTLEEFSPVVERLGFDEFVDLTE 134
 AK CP+L+++ + RYR S + +L E++ +VE + DE ++D+T+
 Sbjct: 65 WQAKRHCPQLVLP-PNFDYRNSSRAMFTILREYTDLVEPVSIDEGYMDMTD 116

Score = 137 (20.6 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
 Identities = 43/148 (29%), Positives = 75/148 (50%)

Query: 178 QIAAEMREAMYNQLGLTGCAGVASNKLLAKLVSGVFKPNQQT VLLPESQHLI HSLNHIK 237
 + A E++ + +L L G+A NK LAK+ S + KP T+L ++ L +
 Sbjct: 125 ETKAIQSRQLQKELLPSISIGIAPNKLAKMASDMKKPLGITILRKRQVPDILWPLP-VG 183

Query: 238 EIPGIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSG 297
 E+ G+G KTA+ L+ LGI+++ +L L++ LGI+ R++ + G ++PV
 Sbjct: 184 EMHGVGKTAELKGLGIHTIGELAADEHSLKRLGGIN-GPRLKNKANGIHHAPV---- 238

Query: 298 PPQSFSEEDSFKKCTSEVEAKNKIEELL 325
 P+ E S ++ + EELL

Sbjct: 239 DPERIYEFKSVGNSSTLSHDSSDEEELL 266

Pedant information for DKF2phfbr2_72b18, frame 2

Report for DKF2phfbr2_72b18.2

[LENGTH] 715
 [MW] 80300.63
 [pI] 6.37
 [HOMOL] TREMBL:SPBC16A3_11 gene: "SPBC16A3.11"; product: "hypothetical protein";
 S.pombe chromosome II cosmid c16A3. 5e-30
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
 repair) [S. cerevisiae, YDR419w] 2e-15
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M.
 genitalium, MG360] 3e-13
 [PIRKW] SOS mutagenesis 2e-11
 [PIRKW] DNA repair 2e-11
 [PIRKW] induced mutagenesis 2e-11
 [SUPFAM] umuC protein 3e-29
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 15
 [PROSITE] PROKAR_LIPOPROTEIN 1
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 21
 [PROSITE] ASN_GLYCOSYLATION 5
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 4.20 %

SEQ MELADVGAASSQGVHDQVLTPNASSRVIVHVDLDCFYAQVEMISNPCLKDKPLGVQOK
 SEG
 PRD ccc

SEQ YLVVTCNYEARKLGVKKLMNVDRDAKEKCPQLVLVNGEDLTRYREMSYKVTELEEFSPV
 SEG
 PRD ceeeehh

SEQ ERLGFDENFVDLTEMVEKRLQQLQSDLSAVTVSGHYNNQSIINLLDVLHIRLLVGSQIA
 SEG
 PRD eeeccchhh

SEQ AEMREAMYNQLGLTGACGAVASNKLLAKLVSGVFKPNQQTVLLPESCQHLIHSLNHIKEIP
 SEG
 PRD hhh

SEQ GIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSGPPQ
 SEG
 PRD cchhh

SEQ SFSEEDSFKKCTSEVEAKNKIEELLASLLNRVCQGRKPHTVRLIIRYSSEKHYGRESR
 SEG
 PRD cccccccccchhh

SEQ QCPIPSHVIQKLTGNVDVMTPMVDILMKLFRNMVNVKMPFHLTLLSVCFCNLKALNTAK
 SEG
 PRD cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ KGLIDYYLMPSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDFLPSGRIESTRTRESPLD
 SEG
 PRD hhhheeecc

SEQ TTNFSKEKDINEFPLCSLPEGVDQEVSKQLPVDIQEEILSGKSREKFQKGSVSCPLHAS
 SEG
 PRD cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ RGVLSFFSKQMQDIPINPRDHLSSSKQVSSVSPCEPGTSGFNSSSSSYMSQKDYSYYL
 SEGxx
 PRD hccchhhh

SEQ DNRLKDERISQGPKEPQGFHTNSNPAVSAFHSFPNLQSEQLFSRNHTTDSHKQTVATDS
 SEG
 PRD hhh

SEQ HEGLTENREPDSDVEKITFPSDIDPQVFYELPEAVQKELLAEWKRTGSDFHIGHK
 SEG
 PRD ccc

Prosites for DKFZphfbr2_72b18.2

PS00001	24->28	ASN_GLYCOSYLATION	PDOC00001
PS00001	160->164	ASN_GLYCOSYLATION	PDOC00001
PS00001	483->487	ASN_GLYCOSYLATION	PDOC00001
PS00001	583->587	ASN_GLYCOSYLATION	PDOC00001
PS00001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS00004	309->313	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	347->351	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	106->109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	201->204	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	307->310	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	438->441	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	459->462	PKC_PHOSPHO_SITE	PDOC00005
PS00005	466->469	PKC_PHOSPHO_SITE	PDOC00005
PS00005	471->474	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	565->568	PKC_PHOSPHO_SITE	PDOC00005
PS00005	592->595	PKC_PHOSPHO_SITE	PDOC00005
PS00005	651->654	PKC_PHOSPHO_SITE	PDOC00005
PS00006	46->50	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	301->305	CK2_PHOSPHO_SITE	PDOC00006
PS00006	303->307	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	448->452	CK2_PHOSPHO_SITE	PDOC00006
PS00006	459->463	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	497->501	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	592->596	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	681->685	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	101->108	TYR_PHOSPHO_SITE	PDOC00007
PS00007	348->356	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	192->198	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	663->669	MYRISTYL	PDOC00008
PS00009	335->339	AMIDATION	PDOC00009
PS00013	186->197	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_72b18.2)

DKFZphfbr2_72d13

group: brain derived

DKFZphfbr2_72d13 encodes a novel 165 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

seems to be testis specific 9 of 10 EST hits are from testis libraries

Sequenced by LMU

Locus: unknown

Insert length: 723 bp

Poly A stretch at pos. 704, no polyadenylation signal found

```

1 AGGGGGGGTA TGGGGGAGGG GGAGACTCTG CAGGAGCCTA ATTCCCCACT
51 CTGAGCTCAC CCTTCTGTCT GCCCGGGGCC TACCCCTTCC CCTACTCTCA
101 CCCTTATAAT CCTTTTCAGC ACTAGGTCTT CCCGTCACCT CCACCTCTCT
151 CCATGACCCG GCTCTGCTTA CCCAGACCCG AAGCACGTGA GGATCCGATC
201 CCAGTTCCCT CAAGGGGCCT GGGTGCTGGG GAGGGGTCAG GTAGTCCAGT
251 GCGTCCACCT GTATCCACCT GGGGCCCTAG CTGGGCCCAG CTCCTGGACA
301 GTGTCTTATG GCTGGGGGCA CTAGGACTGA CAATCCAGGC AGTCTTTTCC
351 ACCACTGGCC CAGCCCTGCT GCTGCTTCTG GTCAGCTTCC TCACCTTTGA
401 CCTGCTCCAT AGGCCCCGAG GTCACACTCT GCCACAGCGC AAACCTTCTCA
451 CCAGGGGGCA GAGTCAGGGG GCCGGTGAAG GTCCTGGACA GCAGGAGGCT
501 CTAATCCTGC AAATGGGTAC AGTCTCAGGA CAACTTAGCC TCCAGGACGC
551 ACTGCTGCTG CTGCTCATGG GGCTGGGCCC GCTCCTGAGA GCCTGTGGCA
601 TGCCCTTGAC CCTGCTTGGC CTGGCTTCTT GCCTCCATCC TTGGGCCTGA
651 GAGCCCTTCC CCACAACTCA GTGTCCTTCA AATATACAAT GACCACCCTT
701 CTTCAAAAAA AAAAAAAAAA AAC

```

BLAST Results

Entry HS860F19 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 860F19
 Score = 2059, P = 1.1e-85, identities = 423/434
 2 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 153 bp to 647 bp; peptide length: 165
 Category: putative protein
 Classification: no clue

```

1 MTRLCLPRPE AREDPPIVPP RGLGAGEGSG SPVRPPVSTW GPSWAQLLDS
51 VLWLGLALGLT IQAVFSTTGP ALLLLLVSL TFDLLHRPAG HTLPQRKLLT
101 RGQSQGAGEG PGQEQALLLO MGTVSGQLSL QDALLLLMG LGPLLRLACGM
151 PLTLLGLAFC LHPWA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72d13, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_72d13, frame 3

Report for DKFZphfbr2_72d13.3

```
[LENGTH]      165
[MW]           17393.73
[pI]           7.80
[BLOCKS]       BL00068A Malate dehydrogenase proteins
[KW]           TRANSMEMBRANE 2
[KW]           LOW COMPLEXITY 29.70 %
```

[illegible]

```
SEQ      IQAVFSTGTPALLLLLVSFITFDLLHRPAGHTLPQRKLLTRGSQGAGEPGQQEALLLQ
SEG      ..... xxxxxxxxxxxxxxxx ..... xxxxxxxxxxxxxxxx ....
PRD      eeeeeccccchhhhhhhhhhhhhcccccceccccccccccccccccccccchhhhhh
MEM      MMMMMmmmmmmmmmmmmmmmmmm
```

```

SEQ      MGTVSGQLSLQDALLLLMLGLPLLRACGMPLTLLGLAFLHPWA
SEG      . . . . .XXXXXXXXXXXXXXXXXXXXX. . . . .
PRD      hccccchhhhhhhhhhhhhhhccccchhhhhhhcccccc
MEM      MTTMMTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

```

(No Prosite data available for DKF2phfbr2 72d13.3)

(No Pfam data available for DKFZphfbr2_72d13.3)

DKFZphfbr2_72112

group: nucleic acid management

Summary DKFZphfbr2_72112 encodes a novel 344 amino acid protein with similarity to YDR126w and other *S. cerevisiae* proteins.

The novel protein contains a myc-type, helix-loop-helix dimerization domain signature. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, the protein could be a novel DNA-binding protein.

The new protein can application in modulating gene expression.

similarity to YDR126w ;
membrane regions: 2

similarity to YDR126w

complete cDNA complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1270 bp
Poly A stretch at pos. 1251, no polyadenylation signal found

```
1 GGGGGCGCCC GGGAGGCGCC GGAGCCCAGC GGCTGGCGCC AGATCCAGGC
51 TCCTGGAAGA ACCATGTCCG GCAGCTACTG GTCATGCCAG GCACACACTG
101 CTGCCCAGA GAGGCTGCTG TTTGAATTAT CTGTGAATGT TGGGAAGAGG
151 AATGCCAGAG CTGCCGCTG AAAATTACCC AACCAAGAGA AATCTGCAGG
201 ATGGACTTTC TGGTCCTCTT CTTGTTCTAC CTGGCTTCGG TGCTGATGGG
251 TCTTGTCTT ATCTGCGTCT GCTCGAAAAC CCATAGCTTG AAAGGCCTGG
301 CCAGGGGAGG AGCACAGATA TTTTCCTGTA TAATTCCAGA ATGTCTTCAG
351 AGAGCCGTGC ATGGATTGCT TCATTACCTT TTCCATACGA GAAACCACAC
401 CTTCAATTGT CTGCACCTGG TCTTGCAAGG GATGGTTTAT ACTGAGTACA
451 CCTGGGAAGT ATTTGGCTAC TGTCAGGAGC TGGAGTTGTC CTTGCATTAC
501 CTTCTTCTGC CCTATCTGCT GCTAGGTGTA AACCTGTTT TTTTCACCCT
551 GACTTGTGGA ACCAATCCTG GCATTATAAC AAAAGCAAAT GAATTATTAT
601 TTCTTCAATG TTATGAATTT GATGAAGTGA TGTTTCCAAA GAACGTGAGG
651 TGCTCTACTT GTGATTTAAG GAAACCAGCT CGATCCAAGC ACTGCACTGT
701 GTGTAATCGG TGTGTGCACC GTTTCGACCA TCACTGTGTT TGGGTGAACA
751 ACTGCATCGG GGCCTGGAAC ATCAGGTAAT TCCTCATCTA CGTCTTGACC
801 TTGACGGCCT CGGCTGCCAC CGTCGCCATT GTGAGCACC CTTTCTGGT
851 CCACTTGGTG GTGATGTCAG ATTTATACCA GGAGACTTAC ATCGATGACC
901 TTGGACACCT CCATGTTATG GACACGGTCA TTCTTATTCA GTACCTGTTT
951 CTGACTTTTC CACGGATTGT CTTATGCTG GGCTTTGTCG TGGTCCTGAG
1001 CTTCTCTCTG GGTGGCTACC TGTGTCTGT CTTGTATCTG GCGGCCACCA
1051 ACCAGACTAC TAACGAGTGG TACAGAGGTG TCTGGGCCTG GTGCCAGCGT
1101 TGCCCTTGGT TGGCCTGGCC TCCGTCAGCA GAGCCCCAAG TCCACCGGAA
1151 CATTTCACTC CATGGGCTTC GGAGCAACCT TCAAGAGATC TTTCTACCTG
1201 CCTTTCCATG TCATGAGAGG AAGAAACAAG AATGACAAGT GTATGACTGC
1251 CAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1232 bp; peptide length: 344
Category: similarity to unknown protein

```

1  M D F L V L F L F Y  L A S V L M G L V L  I C V C S K T H S L  K G L A R G G A Q I  F S C I I P E C L Q
51 R A V H G L L H Y L  F H T R N H T F I V  L H L V L Q G M V Y  T E Y T W E V F G Y  C Q E L E S L S H Y
101 L L L P Y L L L G V  N L F F F T L T C G  T N P G I I T K A N  E L L F L H V Y E F  D E V M F P K N V R
151 C S T C D L R K P A  R S K H C S V C N W  C V H R F D H H C V  W V N N C I G A W N  I R Y F L I Y V L T
201 L T A S A A T V A I  V S T T F L V H L V  V M S D L Y Q E T Y  I D D L G H L H V M  D T V I L I Q Y L F
251 L T F P R I V F M L  G F V V V L S F L L  G G Y L L S V L Y L  A A T N Q T T N E W  Y R G V W A W C Q R
301 C P L V A W P P S A  E P Q V H R N I H S  H G L R S N L Q E I  F L P A F F C H E R  K K Q E

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72112, frame 3

TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein";
S.pombe chromosome II cosmid c13G1., N = 2, Score = 247, P = 1.4e-22

TREMBL:CED2021_3 gene: "D2021.2"; Caenorhabditis elegans cosmid
D2021., N = 1, Score = 209, P = 9e-17

TREMBL:CEC43H6_2 gene: "C43H6.7"; Caenorhabditis elegans cosmid
C43H6., N = 1, Score = 206, P = 5.2e-15

PIR:S52691 probable membrane protein YDR126w - yeast (Saccharomyces
cerevisiae), N = 1, Score = 207, P = 8.4e-15

PIR:E71607 metal binding protein (DHHC domain) PFB0725c - malaria
parasite (Plasmodium falciparum), N = 1, Score = 182, P = 1.1e-13

>TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein";
S.pombe chromosome II cosmid c13G1.
Length = 356

HSPs:

Score = 247 (37.1 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
Identities = 55/148 (37%), Positives = 85/148 (57%)

```

Query:   52  A V H G L L H Y L F H T R N H -- T F I V L H L V L Q G M --- V Y T E Y T W E V F G Y C Q E L E S L S H Y L L L P Y 105
          A+  L +Y+ +  N    F+ L L+  G+    +Y    + F    + + L  +LLPY
Sbjct:   64  A M R S L S N Y V L Y K N N P L V V F L Y L A L I T I G I A S F F I Y G S S L T Q K F S I I D W I S V - L T S V L L P Y 122

```

```

Query:   106 L L L G V N L F F F T L T C G T N P G I I T K A N E L L F L H V Y E F D - E V M F P K N V R C S T C D L R K P A R S K H 164
          ++L+  +  +NPG I  N    + +D ++ FP  +CSTC  KPARSKH
Sbjct:   123 ---ISLY---IAAKSNPGKIDLKNWNEASRRFPYDYKIFFPN--KCSTCKFEKPARSKH 173

```

```

Query:   165 C S V C N W C V H R F D H H C V W V N N C I G A W N I R Y F L I Y V L 199
          C +CN CV +FDHHC+W+NNC+G  N RYF +++L
Sbjct:   174 C R L C N I C V E K F D H H C I W I N N C V G L N N A R Y F F L F L L 208

```

Score = 43 (6.5 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
Identities = 10/35 (28%), Positives = 17/35 (48%)

```

Query:   257 V F M L G F V V - V L S F L L G G Y L L S V L Y L A A T N Q T T N E W 290
          V F ++ +  V L  L G Y  ++Y  T  + +W
Sbjct:   254 V F L I S L I C S V L V L C L L G Y E F F L V Y A G Y T T N E S E K W 288

```

Pedant information for DKFZphfbr2_72112, frame 3

Report for DKFZphfbr2_72112.3

```

[LENGTH]      344
[MW]           39677.23
[pI]           7.26
[HOMOL]        TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein"; S.pombe
chromosome II cosmid c13G1. 3e-17
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YDR126w] 1e-16
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YDR264c] 8e-05
[FUNCAT]       10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]
8e-05
[PIRKW]        transmembrane protein 4e-15
[SUPFAM]        ankyrin repeat homology 1e-10
[SUPFAM]        unassigned ankyrin repeat proteins 1e-10
[PROSITE]       MYRISTYL 4
[PROSITE]       CK2_PHOSPHO_SITE 3

```

```

[PROSITE]      PKC_PHOSPHO_SITE      1
[PROSITE]      ASN_GLYCOSYLATION     2
[KW]           SIGNAL_PEPTIDE 30
[KW]           TRANSMEMBRANE  2
[KW]           LOW_COMPLEXITY      16.57 %

SEQ      MDFLVLFLFYLASVLMGLVLICVCSKTHSLKGLARGGAQIFSCIIPECLQRAVHGLLHYL
SEG      .....
PRD      ccchhhhhhhhhhhhhheeeeeccccceeeeeccccceeeeeehhhhhhhhhhhheee
MEM      .....

SEQ      FHTRNHTFIVLHLVLQGMVYTEYTWEVFGYCQELSLHYLLPYLLGVNLFFFTLTCG
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      eccccchhhhhhhhhccchhhhhhhheeeccccceehhhhhhhhhhhhhccceeeccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      TNPGIITKANELLFLHVIYEFDEVMFPPKNVRCSTCDLRKPARSKHCSVCNWCVHRFDHHCV
SEG      .....
PRD      cccccccccchhhhhhhhhcccccccccecccccccccccccccccccccccccccccc
MEM      M.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      WVNNCIGAWNIRYFLIYVLTLTASAATVAIVSTTFLVHLVVMSDLYQETYIDDLGHLHVM
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccchhhhhhhhhccchhhhhhhhhhhhhhhhhhhccccccccccccccccchh
MEM      .....

SEQ      DTVILIQLYFLTFFPRIVFMLGFVVVLSFLGGYLLSVLYLAATNQTTNEWYRGVWAWCQR
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      hhhhhhhhhhhhhhhccccccccceccccchhhhhhhhhccchhhhhhhhhhhccccc
MEM      .....

SEQ      CPLVAWPSPAEPQVHRNIHSHGLRSNLQEIFLPAFPCHERKKQE
SEG      .....
PRD      ccccccccccccccecccccccccccccecccccccccccccc
MEM      .....

```

Prosites for DKFZphfbr2_72112.3

PS00001	65->69	ASN_GLYCOSYLATION	PDOC00001
PS00001	284->288	ASN_GLYCOSYLATION	PDOC00001
PS00005	29->32	PKC_PHOSPHO_SITE	PDOC00005
PS00006	152->156	CK2_PHOSPHO_SITE	PDOC00006
PS00006	229->233	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	77->83	MYRISTYL	PDOC00008
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	322->328	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_72112.3)

DKFZphfbr2_72m16

group: unknown

DKFZphfbr2_72m16 encodes a novel 287 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 1462 bp

Poly A stretch at pos. 1441, polyadenylation signal at pos. 1421

```

1 GGGGAGGACC GGAGGACCGA GGACAGAAAG ATTGGTGGAC AGGAGCAGCG
51 GCCGGTGGGG AGGGCGCTCG GCGGCGGCCT GCGGCCATGG CCACCGTGAT
101 GGCAGCGACG GCGGCGGAGC GGGCGGTGCT GGAGGAGGAG TTCCGCTGGC
151 TGCTGCACGA CGAGGTGCAC GCTGTGTGA AGCAGCTGCA GGACATCCTC
201 AAGGAGGCCT CTCTGCGCTT CACTCTGCCG GGCTCCGGCA CTGAGGGGCC
251 CGCCAAGCAA GAGAACTCA TCCTAGGCAG CTGTGGCACA GACCAGGTGA
301 AGGGTGTGCT GACTCTGCAG GGGGATGCCC TCAGCCAGGC GGATGTGAAC
351 CTGAAGATGC CCGGAACAA CCAGCTGCTG CACTTCGCCT TCCGGGAGGA
401 CAAGCAGTGG AAGCTGCAGC AGATCCAGGA TGCCAGAAAC CATGTGAGCC
451 AAGCCATTTA CCTGCTTACC AGCCGGGACC AGAGCTACCA GTTCAAGACG
501 GGGCGCTAGG TCCTCAAGCT GATGGACGCA GTGATGCTGC AGCTGACCAG
551 AGCCCGAAAC CGGCTCACC ACCCCGCCAC CCTCACCCTC CCGGAGATCG
601 CCGCCAGCGG CCTCACGCGG ATGTTGCGCC CTGCCCTGCC GTCCGACCTG
651 CTGGTCAACG TCTACATCAA CCTCAACAAG CTCTGCCTCA CGGTGTACCA
701 GCTGCATGCC CTGCAGCCCA ACTCCACCAA GAACTTCCGC CCAGCTGGGG
751 GCGCGGTGCT GCATAGCCCT GGGGCCATGT TCGAGTGGGG CTCTCAGCGC
801 CTGGAGGTGA GCCACGTGCA CAAAGTGGAG TCGGTGATCC CCTGGCTCAA
851 CGACGCCCTG GTCTACTTCA CCGTCTCCCT GCAGCTCTGC CAGCAGCTTA
901 AGGACAAGAT CTCCGTGTTT TCCAGTACT GGAGCTACAG ACCCTTCTGA
951 TCACAGCACC CAGGAGCTTG TCTCCAGGAA GCGGCCCCCG TCCCTTACTC
1001 ATACCCACCA CAGAGCACC A GCCAGTGCCA ACGCCAGGCT GCTATTTATC
1051 TCCTATCCCA ACCCCCTACC CCACCTAACA CATTGCACT GCCGGGAATG
1101 GACACTGGAA GTGCCAGGAG GAAGGAAGGC TGGTTTGGTG GGGTAGTGGG
1151 GAGGTCAAGG AGGCGGGGCC AAGGGTGTCC CACATTCCCA ACACCGCCCT
1201 CTGATACCA TGGGAATCTT TGGACTCAGG ACAGGGCCAG GCGCAGGGCT
1251 CTCCCTCCTC TCCCTTCCG TGTCCCTCC CCCTGGAGGG CATGGTGTCTG
1301 GGGGGTGGCA CTGAGCTATG AGTCCCAGGG ATGGTGAGGA ACGCCACAGA
1351 CAGAGCCACC CTAGGAGTGA GTATAGTGCT GGTGACTGTG TTTCATAGCC
1401 CCAGTCCAGG GCTGTCTAAG AAATAAGAT CATCAGACT CAAAAA
1451 AAAAAA AC

```

BLAST Results

Entry HS604351 from database EMBL:

human STS WI-18474.

Score = 1178, P = 1.5e-48, identities = 250/268

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 87 bp to 947 bp; peptide length: 287
 Category: similarity to unknown protein

1	MATVMAATAA	ERAVLEEFR	WLLHDEVHAV	LKQLQDILKE	ASLRFTLPGS
51	GTEGPAKQEN	FILGSCDGTQ	VKGVLTQSD	ALSADPNVLK	MPRNNQLLHV
101	AFREDQKOWL	QQIQDARNHW	SQAIYLLTGR	DQSYQFKTAL	EVLKMLDMVM
151	LQLTRARNRL	TTPATLTLPE	IAASGLTRMF	APALPSDLV	NVYINLNKLC
201	LTYYQLHALQ	PNSTKNFRPA	GGAVLSHPSG	MEFWGSQRLE	VSHVHKVECV
251	IPWLNDALQY	FVTSLCOFCA	LKDKISVSSA	YWSYRFF	

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2 72ml6, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 72m16, frame 3

Report for DKFZphfbr2 72m16.3

```
[LENGTH]      287
[MW]           32254.40
[pI]           8.30
[HOMOL]       TREMBL:AF025459 2 gene: "H14A12.3"; Caenorhabditis elegans cosmid H14A12. 3e-14
```

[PROSITE]	MYRISTYL	1
[PROSITE]	CK2_PHOSPHO_SITE	6
[PROSITE]	PKC_PHOSPHO_SITE	5
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha_Beta	
[KW]	LOW_COMPLEXITY	6.27 %

[illegible]

Prosite for DKFZphfbr2 72m16.3

PS000001	212->216	ASN_GLYCOSYLATION	PDOC000001
PS000005	42->45	PKC_PHOSPHO_SITE	PDOC000005
PS000005	128->131	PKC_PHOSPHO_SITE	PDOC000005
PS000005	213->216	PKC_PHOSPHO_SITE	PDOC000005
PS000005	236->239	PKC_PHOSPHO_SITE	PDOC000005
PS000005	283->286	PKC_PHOSPHO_SITE	PDOC000005
PS000006	8->12	CK2_PHOSPHO_SITE	PDOC000006
PS000006	50->54	CK2_PHOSPHO_SITE	PDOC000006
PS000006	83->87	CK2_PHOSPHO_SITE	PDOC000006
PS000006	128->132	CK2_PHOSPHO_SITE	PDOC000006
PS000006	138->142	CK2_PHOSPHO_SITE	PDOC000006
PS000006	167->171	CK2_PHOSPHO_SITE	PDOC000006
PS000008	64->70	MYRISTYL	PDOC000008

(No Pfam data available for DKF2phfbr2 72m16.3)

DKFZphfbr2_72n12

group: brain derived

DKFZphfbr2_72n12 encodes a novel 117 amino acid protein with similarity to a protein with conserved sequence in bacteria and eukariota.

The novel protein is very similar to human MM46, human and rat ganglioside expression factor-2 (GEF2), *C. elegans* 14.8 kD protein C32D5.9 and *Laccaria bicolor* symbiosis-related protein LBU93506_1. The function of this highly conserved proteins is not known.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to rat GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="12"

Insert length: 1880 bp

Poly A stretch at pos. 1859, polyadenylation signal at pos. 1830

```
1 GGGGGCGCGT ATTTCTCCAT CTGGCTCTCC TCTACCTCCA GGCAGGCTCA
51 CCCGAGATCC CGCCCCCGAA CCCCCCTGCG ACACCTCGGC CAGCGCTGTT
101 GCCCCCGGAG CGGACGTTTC TGCAGCTATT CTGAGCACAC CTGACGTCG
151 GCTGAGGGAG CGGGACAGGG TCAGCGGCGA AGGAGGCAGG CCCC CGCGCG
201 GGATCTCGGA AGCCCTGCGG TGCATCATGA AGTTCAGTA CAAGGAGGAC
251 CATCCCTTTG AGTATCGGAA AAAGGAAGGA GAAAAGATCC GGAAGAAATA
301 TCCCGACAGG GTCCCGGTGA TTGTAGAGAA GGCTCCAAA GCCAGGGTGC
351 CTGATCTGGA CAAGAGGAAG TACCTAGTGC CCTCTGACCT TACTGTTGGC
401 CAGTTCTACT TCTTAATCCG GAAGAGAATC CACCTGAGAC CTGAGGACGC
451 CTTATTCTTC TTTGTCAACA ACACCATCCC TCCCACCACT GCTACCATGG
501 GCCAACTGTA TGAGGACAAT CATGAGGAAG ACTATTTTCT GTATGTGGCC
551 TACAGTGATC AGAGTGTCTA TGGGAAATGA GTGGTTGGAA GCCCAGCAGA
601 TGGGAGCACC TGGACTTGGG GGTAGGGGAG GGGTGTGTGT GCGCGACATG
651 GGGAAAGAGG GTGGCTCCCA CCGCAAGGAG ACAGAAAGTG AAGACATCTA
701 GAAACATTAC ACCACACACA CCGTCATCAC ATTTTCACAT GCTCAATTGA
751 TATTTTTTGC TGCTTCCTCG GCCCAGGGAG AAAGCATGTC AGGACAGAGC
801 TGTTGGATTG GCTTTGATAG AGGAATGGGG ATGATGTAAG TTTACAGTAT
851 TCCTGGGGTT TAATTGTTGT GCAGTTTCAT AGATGGGTCA GGAGGTGGAC
901 AAGTTGGGGC CAGAGATGAT GGCAGTCCAG CAGCAACTCC CTGTGCTCCC
951 TTCTCTTTGG GCAGAGATTC TATTTTGTAC ATTTGCACAA GACAGGTAGG
1001 GAAAGGGGAC TTGTGGTAGT GGACCATACC TGGGGACCAA AAGAGACCCA
1051 CTGTAATTGA TGCATTGTGG CCCCTGATCT TCCCTGTCTC ACACTTCTTT
1101 TCTCCCATCC CGGTTGCAAT CTCACTCAGA CATCACAGTA CCACCCAGG
1151 GGTGGCAGTA GACAACAACC CAGAAATTTA GACAGGGATC TCTTACCTTT
1201 GGAATATAGG GGTAGGCAT GAAGGTGGTT GTGATTAAAG AGATGGTTTT
1251 GTTATTAAAT AGCATTAAAC TGGAATTGAC AAGAGTGTG AGCATCCCTG
1301 TCTAACCTGC TCTTTCTCTT TGGTGCCCTT TATCTCACCC CTTCCTTGGG
1351 ATTTAATAAG TCTCAGGCAT TTCCAATTGT AGACTAAAAC CACTCTTAGC
1401 ATCTCCTCTA GTATTTTCCA TGTATCAGGA AAGAGGTGTC TTATGTAGGG
1451 AGGGGGCAAG TATGAAGTAA GGTAATTATA TACTACTCTC ATTCAGGATT
1501 CTTGCTCCCA TGCTGCTGTC CCTTCAGGCT CACATGCACA GGAATGCTAC
1551 ATGATGGCCA GCTGCTTCCC TCCTTGTTTA TCATCCACTG CAGCTGCTAG
1601 TTAGAAAGGT TTGGAGGGAT GACTTTTGTG AAATCATGGG GATTTTATTG
1651 ATTTATTTTC ACTTTTGGGA TTTTGTGGGG TGGGAGTGGG GAGCAGGAAT
1701 TGCACTCAGA CATGACATT CAATTCATCT CTGCTAATGA AAAGGGTTCT
1751 TTCTCTTGGG GGAATGTGT GTGTCAGTTC TGTGCTGTC AAGTTCTTGT
1801 ATAATGAAGT CAATGCCATC AGGCCAAGGA AATAAAATAA TTGCTTACCT
1851 TAAAAATCGA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry HS418210 from database EMBL:

human STS SHGC-10496.

Score = 1916, P = 4.0e-80, identities = 394/400

Entry AC006514 from database EMBLNEW:

*** SEQUENCING IN PROGRESS *** Homo sapiens; HTGS phase 1, 68 unordered pieces.

Score = 610, P = 2.7e-16, identities = 128/134

4 exons

Medline entries

 No Medline entry

Peptide information for frame 2

 ORF from 227 bp to 577 bp; peptide length: 117
 Category: strong similarity to known protein

1 MKFQYKEDHP FEYRKKEGK IRKKYPDRVP VIVEKAPKAR VPDLDKRRKYL
 51 VPSDLTVGQF YFLIRKRIHL RPEDALFFV NNTIPPTSAT MGQLYEDNHE
 101 EDYFLYVAYS DESVYGK

BLASTP hits

Entry YQD9_CAEEL from database SWISSPROT:
 HYPOTHETICAL 14.8 KD PROTEIN C32D5.9 IN CHROMOSOME II.
 Score = 496, P = 1.8e-47, identities = 91/116, positives = 105/116

Entry SYRP_LACBI from database SWISSPROT:
 SYMBIOSIS-RELATED PROTEIN.
 Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry LBU93506_1 from database TREMBL:
 product: "symbiosis-related protein"; Laccaria bicolor
 symbiosis-related protein mRNA, partial cds.
 Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry GEF2_RAT from database SWISSPROT:
 GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2).
 Score = 373, P = 2.0e-34, identities = 71/116, positives = 88/116

Alert BLASTP hits for DKFZphfbr2_72n12, frame 2

TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete
 cds., N = 1, Score = 549, P = 4.7e-53

SWISSPROT:GEF2_HUMAN GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)., N = 1,
 Score = 373, P = 2.1e-34

>TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete
 cds.
 Length = 117

HSPs:

Score = 549 (82.4 bits), Expect = 4.7e-53, P = 4.7e-53
 Identities = 101/116 (87%), Positives = 110/116 (94%)

Query: 1 MKFQYKEDHPFEYRKKEGKIRKKYPDRVPVIVEKAPKARVPDLKRRKYLVPDLTVGQF 60
 MKF YKE+HPFE R+ EGEKIRKKYPDRVPVIVEKAPKAR+ DLDK+KYLVPDLTVGQF
 Sbjct: 1 MKFVYKEEHPFEKRRSEGEKIRKKYPDRVPVIVEKAPKARIGDLKKKYLVPDLTVGQF 60
 Query: 61 YFLIRKRIHLRPEDALFFVNNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYG 116
 YFLIRKRIHLR EDALFFVNN IPPTSATMGQLY+++HEED+FLY+AYSDESVYG
 Sbjct: 61 YFLIRKRIHLRAEDALFFVNNVIPPTSATMGQLYQEHHEEDFFLYIAYSDESVYG 116

Pedant information for DKFZphfbr2_72n12, frame 2

Report for DKFZphfbr2_72n12.2

[LENGTH] 117
 [MW] 14044.07
 [pI] 8.67
 [HOMOL] TREMBL:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete cds. 1e-56

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBL078c] 4e-36
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YBL078c] 4e-36
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YBL078c] 4e-36
[SUPFAM] hypothetical protein YBL078c 8e-35
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta

SEQ MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDRKYLVPSDLTVGQF
PRD cccccccccchhhhhhhhhhhhhhhccccceeeccccccccccccceccccchhhh

SEQ YFLIRKRIHLRPEDALFFVNNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVMYK
PRD hhhhhhhhhccccceeeccccccccchhhhhhhhhccccceeecccccccccc

Prosites for DKFZphfbr2_72n12.2

PS00001 81->85 ASN_GLYCOSYLATION PDOC00001

(No Pfam data available for DKFZphfbr2_72n12.2)

DKFZphfbr2_78c24

group: signal transduction

DKFZphfbr2_78c24 encodes a novel 563 amino acid protein with strong similarity to guanylate-binding proteins (GBPs).

GBPs were originally described as proteins that are strongly induced by interferons and are capable of binding to agarose-immobilized guanine nucleotides. hGBP1, the first of two members of this protein family in humans, represents a novel type of GTPase. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and a RGD cell attachment site. It seems to be a new member of the GBP-family and shows a splicing pattern not described previously.

The new protein can find application in modulating/blocking the response of cells to interferons.

strong similarity to guanine nucleotide-binding protein 1/2
but different "splice variant" aa 211-245 of GBP1/2 missing

Sequenced by MediGenomix

Locus: unknown

Insert length: 2952 bp

Poly A stretch at pos. 2927, polyadenylation signal at pos. 2914

```
1 CAGTTTCATT AGGCTCTGAA GCCATTACAA AGGTTGCTTA ACTTCTAATT
51 ATTTGATCAC TGAGGAAAAT CCAGAAAGCT ACACAACACT GAAGGGGTGA
101 AATAAAAGTC CAGCGATCCA GCGAAAGAAA AGAGAAGTGA CAGAAACAAC
151 TTTACCTGGA CTGAAGATAA AAGCACAGAC AAGAGAACAA TGCCTCGGAC
201 ATGGCTCCAG AGATCCACAT GACAGGCCCA ATGTGCCTCA TTGAGAACAC
251 TAATGGGGAA CTGGTGGCGA ATCCAGAAGC TCTGAAAATC CTGTCTGCCA
301 TTACACAGCC TGTGGTGGTG GTGGCAATTG TGGGCCTCTA CCGCACAGGA
351 AAATCCTACC TGATGAACAA GCTAGCTGGG AAGAATAAGG GCTTCTCTCT
401 GGGCTCCACA GTGAAATCTC ACACCAAAGG AATCTGGATG TGGTGTGTGC
451 CTCACCCCAA AAAGCCAGAA CACACCTTAG TCCTGCTTGA CACTGAGGGC
501 CTGGGAGATG TAAAGAAGGG TGACAACCAG AATGACTCCT GGATCTTCAC
551 CCTGGCCGTC CTCCTGAGCA GCACTCTCGT GTACAATAGC ATGGGAACCA
601 TCAACCAGCA GGCTATGGAC CAACTGTACT ATGTGACAGA GCTGACACAT
651 CGAATCCGAT CAAAATCCTC ACCTGATGAG AATGAGAATG AGGATTTCAGC
701 TGACTTTGTG AGCTTCTTCC CAGATTTTGT GTGGACACTG AGAGATTTCCT
751 CCCTGGACTT GGAAGCAGAT GGACAACCCC TCACACCAGA TGAGTACCTG
801 GAGTATTCCC TGAAGCTAAC GCAAGGTAAC AGGAAGCTTG CCGAGCTTGA
851 GAAACTACAA GATGAAGAGC TGGACCTGTA ATTTGTGCAA CAAGTAGCAG
901 ACTTCTGTTC CTACATCTTT AGCAATTCCA AAATAAAAC TCTTTCAGGA
951 GGCATCAAGT TCAATGGGCC TTGTCTAGAG AGCCTAGTGC TGACCTATAT
1001 CAATGCTATC AGCAGAGGGG ATCTGCCCTG CATGGAGAAC GCAGTCTTGG
1051 CCTTGGCCCA GATAGAGAAC TCAGCCGCGAG TGCAAAGGC TATTGCCCCAC
1101 TATGACCAAG AGATGGGCCA GAAGGTGCAG CTGCCCCGAG AAACCTTCCA
1151 GGAGCTGCTG GACCTGCACA GGGTTAGTGA GAGGGAGGCC ACTGAAGTCT
1201 ATATGAAGAA CTCTTTCAAG GATGTGGACC ATCTGTTTCA AAAGAAATTA
1251 GCGGCCCAAG TAGACAAAAA GCGGGATGAC TTTTGTAAAC AGAATCAAGA
1301 AGCATACATCA GATCGTTGCT CAGCTTTACT TCAGGTCAAT TTCAGTCTCT
1351 TAGAAGAAGA AGTGAAGGCG GGAATTTATT CGAAACCAGG GGGCTATTGT
1401 CTCTTTATTCT AGAAGCTACA AGACCTGGAG AAAAGTACT ATGAGGAACC
1451 AAGGAAGGGG ATACAGGCTG AAGAGATTCT GCAGACATAC TTGAAATCCA
1501 AGGAGTCTGT GACCGATGCA ATTCTACAGA CAGACCAGAT TCTCACAGAA
1551 AAGGAAAAGG AGATTGAAGT GGAATGTGTA AAAGTGAAT CTGCACAGGC
1601 TTCAGCAAAA ATGGTGGAGG AAATGCAAAAT AAAGTATCAG CAGATGATGG
1651 AAGAGAAAGA GAAGAGTTAT CAAGAACATG TGAAACAATT GACTGAGAAG
1701 ATGGAGAGGG AGAGGGCCCA CTTGCTGGAA GAGCAAGAGA AGACCCTCAC
1751 TAGTAAACTT CAGGAACAGG CCCGAGTACT AAAGGAGAGA TGCCAAGGTG
1801 AAAGTACCCA ACTTCAAAAT GAGATACAAA AGCTACAGAA GACCCTGAAA
1851 AAAAAACCA AGAGATATAT GTCGCATAAG CTAAAGATCT AAACAACAGA
1901 GCTTTTCTGT CATCCTAACC CAAGGCATAA CTGAACAAT TTTAGAATTT
1951 GGAACCAAGT TCACTATATT TGATAATAAT TAGATCTTGC ATCATACAC
2001 TAAAAGTTTA CAAGAACATG CAGTTCAATG ATCAAAATCA TGTTTTTTCC
2051 TTAAAAAGAT TGTAAATTGT GCAACAAAGA TGCATTACC TCTGTACCAA
2101 CAGAGGAGGG ATCATGAGTT GCCACCACTC AGAAGTTTAT TCTTCCAGAC
2151 GACCAAGTGA TACTGAGGAA AGTCTTAGGT AAAATCTTGG GGACATATTT
2201 GGGCACTGGT TTGGCCAAGT GTACAATAGG TCCCAATATC AGAAACAACC
2251 ATCCTAGCTT CTTAGGGAAG ACAGTGTACA GTTCTCCATT ATATCAAGGC
2301 TACAAGGTCT ATGAGCAATA ATGTGATTTT TGGACATTGC CCATGGATAA
2351 TTCTCACTGA TGGATCTCAA GCTAAAGCAA ACCATCTTAT ACAGAGATCT
2401 AGAATCTTAT ATTTCCATA GGAAGGTAAA GAAATCATT GCAAGAGTAG
2451 GAATTGAATC ATAAACAAAT TGGCTAATGA AGAAATCTTT TCTTCTTGT
2501 TCAATTTCAT TAGATTATAA CCTTAATGTG ACACCTGAGA CCTTTAGACA
```

```

2551 GTTGACCCCTG AATTAAATAG TCACATGGTA ACAATTATGC ACTGTGTAAT
2601 TTTAGTAATG TATAACATGC AATGATGCAC TTAACTGAA GATAGAGACT
2651 ATGTTAGAAA ATTGAACTAA TTTAATTATT TGATTGTTTT AATCCTAAAG
2701 CATAAGTTAG TCTTTTCCTG ATCTTAAAG GTCATACTTG AAATCCTGCC
2751 AATTTTCCCC AAAGGGAATA TGGAAATTTT TTTGACTTTC TTTTGAGCAA
2801 TAAAATAATT GTCTTGCCAT TACTTAGTAT ATGTAGACTT CATCCCAATT
2851 GTCAAACATC CTAGGTAAAG GGTGACATT TCTTACAGCA ATTACAGATT
2901 ATTTTGAAC TAGAAATAAA CTAACTAGA AACAAAAAAA AAAAAAAA
2951 AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1889 bp; peptide length: 563
 Category: strong similarity to known protein
 Classification: Cell signaling/communication
 Prosite motifs: RGD (272-275)
 ATP_GTP_A (45-53)

```

1 MAPEIHMTGP MCLIENTNGE LVANPEALKI LSAITQPVVV VAIVGLYRTG
51 KSYLMNKLK KNGGSLGST VKSHTKGIWM WCVPHPKPE HTLVLLDTEG
101 LGDVKKGDNQ NDSWIFTLAV LLSSTLVYNS MGTINQQAMD QLYYVTELTH
151 RIRSKSSPDE NENEDSADFV SFFPDFVWTL RDFSLEAD GQPLTPDEYL
201 EYSLKLTQGN RKLQLEKLQ DEELDPEFVQ QVADFCYIF SNSKTKLSG
251 GIKVNGPCLE SLVLTYNINAI SRGDLPCMEN AVLALAQIEN SAAVQKAIH
301 YDQMGQKQV LPAETLQELL DLHRVSEREA TEVYMKNSEK DVDHFLQKKL
351 AAQLDKKRD FCKQNEASS DRCSALLQVI FSPLEEVKA GIYSKPGGYC
401 LFIQQLQDL KKYEEPRKG IQAEELQTY LKSKESVTD ILQTDQILTE
451 KEKEIEVECV KAESAQASAK MVEEMQIKYQ QMEEKEKSY QEHVKQLTEK
501 MERERAQLE EQEKTLSKL QEQARVLKER CQGESTQLQN EIQLKQKTLK
551 KTKRYMSHK LKI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_78c24, frame 3

PIR:A41268 guanine nucleotide-binding protein 1 - human, N = 2, Score = 1306, P = 4.9e-238

PIR:A46459 macrophage-activation gene-1 protein mag-1 - mouse, N = 2, Score = 942, P = 8.9e-184

PIR:S70524 guanine nucleotide-binding protein 2 - human, N = 2, Score = 1131, P = 4.1e-210

TREMBL:AF077007_1 gene: "Gbp2"; product: "interferon-induced guanylate binding protein GBP-2"; Mus musculus interferon-induced guanylate binding protein GBP-2 (Gbp2) mRNA, complete cds., N = 2, Score = 904, P = 1.2e-179

>PIR:A41268 guanine nucleotide-binding protein 1 - human
 Length = 592

HSPs:

Score = 1306 (195.9 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
 Identities = 264/332 (79%), Positives = 288/332 (86%)

Query: 211 RKLQLEKLQDEELDPEFVQVADFCYIFSNSKTKLSGGIKVNGPCLESVLVTYINAI 270
 RKLQLEKLQDEELDPEFVQVADFCYIFSNSKTKLSGGI+VNGP LESVLVTY+NAI
 Sbjct: 245 RKLQLEKLQDEELDPEFVQVADFCYIFSNSKTKLSGGIQVNGPRLESVLVTYVNAI 304

```

Query: 271 SRGDLPCMENAVLALAQIENSAAVQKAI AHYDQMGQKVQLPAETLQELLDLHRVSREA 330
      S GDLPFCMENAVLALAQIENSAAVQKAI AHY+QMGQKVQLP E+LQELLDLHR SEREA
Sbjct: 305 SSGDLPCMENAVLALAQIENSAAVQKAI AHYEQQMGQKVQLPTESLQELLDLHRDSREA 364

Query: 331 TEVYMKNSFKDVDHLFQKKLAAQLDKKRDDFCCKNQEASSDRCSALLQVIFSPLEEVEKA 390
      EV+++SFKDVDHLFQK+LAAQL+KKRDDFCCKNQEASSDRCS LLQVIFSPLEEVEKA
Sbjct: 365 IEVFIRSSFKDVDHLFQKELAAQLEKKRDDFCCKNQEASSDRCSGLLQVIFSPLEEVEKA 424

Query: 391 GIYSKPGGYCLFIQKLQDLLEKKYEEPRKGIQAEEILQTYLKSKESTDAILQTDQILTX 450
      GIYSKPGGY LF+QKLQDL+KKYEEPRKGIQAEEILQTYLKSKEST+DAILQTDQ LT
Sbjct: 425 GIYSKPGGYRLFVQKLQDLKKKYEEPRKGIQAEEILQTYLKSKESTDAILQTDQTLTE 484

Query: 451 XXXXXXXXXXXXXXXSAQASAKMVEEMQIKYQQMMEEKEKSYQEHVKLTDAIKMXXXXXXXXXX 510
      SAQASAKM++EMQ K +QMME+KE+SYQEH+KQLTEKM
Sbjct: 485 KEKEIEVERVKAESAQASAKMLQEMQQRKNEQMMEQKERSYQEHKQLTEKMENDRVQLLK 544

Query: 511 XXXXKTLTSKLQEQRVLKERCQGESTQLQNEI 542
      +TL KLQE Q +LKE Q ES ++NEI
Sbjct: 545 EOERTLALKLOEOELLKEGFQESRIMKNEI 576

```

Score = 1012 (151.8 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
Identities = 194/211 (91%), Positives = 200/211 (94%)

```

Query:      1 MAPEIHMTGPMCLIENGTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLGA 60
            MA EIHMTGPMCLIENGTNG L+ANPEALKILSAITQP+VVVAIVGLYRTGKSYLMNKLGA
Sbjct:      1 MASEIHMTGPMCLIENGTNGRLMANPEALKILSAITQPMVVVAIVGLYRTGKSYLMNKLGA 60

Query:      61 KNGKGSLSGSTVKSHTKGIWMWCVPHPKKPEHTLVLLDTEGLGDVKKGNDQNDSWIFTLAV 120
            K KGFSLSGSTV+SHTKGIWMWCVPHPKKP H LVLLDTEGLGDV+KGDQNDSWIF LAV
Sbjct:      61 KKKGFSLSGSTVQSHTKGIWMWCVPHPKKPGHILVLLDTEGLGDVEKGDQNDSWIFALAV 120

Query:      121 LLSSTLVYNSMGITINQAMQDLYYVTELTHRIRSKSSPDENENE--DSADFVSFFPDFVW 178
            LLSST VYNS+GTINQAMQDLYYVTELTHRIRSKSSPDENENE DSADFVSFFPDFVW
Sbjct:      121 LLSSTFFVNSIGITINQAMQDLYYVTELTHRIRSKSSPDENENEVEDSADFVSFFPDFVW 180

Query:      179 TLRDFSLDLEADGQPLTPDEYLEYSKLKTQG 209
            TLRDFSLDLEADGQPLTPDEYL YSKL +G
Sbjct:      181 TLRDFSLDLEADGQPLTPDEYLTYSKLKKG 211

```

Pedant information for DKFZphfbr2 78c24, frame 3

Report for DKFZphfbr2 78c24.3

```
[LENGTH]          563
[MW]               64127.72
[pI]               5.45
[HOMOL]            PIR:A41268 guanine nucleotide-binding protein 1 - human 0.0
[SUPFAM]            guanine nucleotide-binding protein 1 0.0
[PROSITE]           ATP_GTP_A      1
[PROSITE]           RGD      1
[KW]                TRANSMEMBRANE 1
[KW]                LOW_COMPLEXITY  6.75 %
[KW]                COILED COIL    10.48 %
```

SEQ	MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLGA
SEG	
PRD	ccccccccceeeeccccchhhhhhhhhhhhhhhccceeeeccccchhhhhhhh
COILSMMMMMMMMMMMMMMMM.....
MEM	
SEQ	KNKGFSLGSTVKSHTKGIWMVCVPHPKPEHTLVLLDTEGLGDVKGDNQNSWIFTLAV
SEG	
PRD	ccccccccccccccccceeeecccccccccceeeecccccccccccccchhhhhhhh
COILS
MEM	
SEQ	LLSSTLVNSMGTINQQAMDQLYYVTELTHRIRSKSSPDENENEDSADFVSFFPDFVWTL
SEG	
PRD	hhhhhheeccccchhhhhhhhhhhhhhhhhhhhhccccccccccceeeecceeeeh
COILS
MEM	
SEQ	RDFSLDLEADGQPLTPDEYLEYSCLKTQGNRLAQLEKLQDEELDPEFVQVADFCSYIF
SEG	
PRD	hhhhhhhhccccccccchhhhhhhhhhhhhccchhhhhhhhhhhhhccchhhhhhhhhhhhhc
COILS	

```

MEM .....
SEQ SNSKTKTLGGIKVNGPCLESVLTYINAIISRGDLPCMENAVLALAQIENSAAVQKAI AH
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
MEM .....

SEQ YDQMGQKQVQLPAETLQELDLHRVSEEREATEVYMKNSEFKDVLDFQKKLAAQLDKKRDD
SEG .....
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
MEM .....

SEQ FCKQNQEASSDRCSALLQVIFSPLEEEVKAGIYSKPGGYCLFIQKLQDLEKYYEPRKG
SEG .....
PRD hhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM .....

SEQ IQAEIILQTYLKSKEVTDAILQTDQILTEKEKEIEVECVKAESAQASAKMVEEMQIKYQ
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM .....

SEQ QMMEEEKESYQEHVKQLTEKMERERAQLLEEQKTLTSLQEQARVLKERCQGESTQLQN
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
MEM .....

SEQ EIQLKQTLKKKTKRYMSHKLKI
SEG ..xxxxxxxxxxxxx.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCC.....
MEM .....

```

Prosites for DKFZphfbr2_78c24.3

PS00016	272->275	RGD	PDOC00016
PS00017	45->53	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphfbr2_78c24.3)

DKFZphfbr2_78d13

group: brain derived

DKFZphfbr2_78d13 encodes a novel 259 amino acid protein with similarity to *C. elegans* putative protein from cosmid K08B12.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to *C.elegans* K08B12.3

Sequenced by MediGenomix

Locus: /map="338.4 cR from top of Chr18 linkage group"

Insert length: 2195 bp

Poly A stretch at pos. 2175, polyadenylation signal at pos. 2156

```
1 CGTCCGTCGG GCAGCAGCGG GGCTGTCTAT CCCGGCTGAG GACCCGCGGC
51 CAGTGGCGGT GGCTGGCTTT GCCATTAGCG GGGGCCTTTC CTGAGGACGG
101 CGTACGGAGT GTGGGGAATG AAGGATGGCA GCATGCCGTG CATTAAAAGC
151 TGTTTTGGTA GATCTCAGTG GCACACTTCA CATTGAAGAT GCAGCTGTGC
201 CAGGCCGACA GGAAGCTCTT AAAAGGTTAC GTGGTGCTTC TGTAAATCATT
251 AGGTTTGTGA CCAATACAAC CAAAGAGAGC AAGCAAGACC TGTTAGAAAG
301 GTTGAGAAAA TTGGAATTTG ATATCTCTGA AGATGAAATA TTCACATCTC
351 TGACTGCAGC CAGAAGTTTA CTAGAGCGGA AACAAGTCAG ACCCATGCTG
401 CTAGTTGATG ATCGGGCACT ACCTGATTTT AAAGGAATAC AAACAAGTGA
451 TCCTAATGCT GTGGTCATGG GATTGGCACC AGAACATTTT CATTATCAAA
501 TTCTGAATCA AGCATTCCGG TTACTCCTGG ATGGAGCACC TCTGATAGCA
551 ATCCACAAAG CCAGGTATTA CAAGAGGAAA GATGGCTTAG CCCTGGGGCC
601 TGGACCATTT GTGACTGCTT TAGAGTATGC CACAGATACC AAAGCCACAG
651 TCGTGGGGAA ACCAGAGAAG ACGTTCTTTT TGAAGCATT GCGGGGCACT
701 GGCTGTGAAC CTGAGGAGGC TGTCATGATA GGAGATGATT GCAGGGATGA
751 TGTGTGTGGG GCTCAAGATG TCGGCATGCT GGGCATCTTA GTAAAGACTG
801 GGAAAAATCG AGCATCAGAT GAAGAAAAAA TTAATCCACC TCCTTACTTA
851 ACTTGTGAGA GTTCCCTCA TGCTGTGGAC CACATTCTGC AGCACCTATT
901 GTGAAGCAAT GTGTGCATCT GAAGCAACTT GAAATGCAGC TTCTTATTGT
951 CTGGAATGAA TCCCTTACCA ACTCAGTGCC AGCATCGGTA GACACCAGTC
1001 AGTGCTGATC GCTTTTAAAC CCTCTTTTGT TGTGCATTAA TTAGAAAGAA
1051 AGGTATTGAA TTGCGGCTAG CCAGTAAGCC TTGCTAATCT CTTTATTTT
1101 GTAACGAAG ATGAGACCCA AAGAAAGGGA AAGCTGAGAT TTTGTGCCAT
1151 TCCTTTTAAA ATATTCATCA GGTAGGTGG GGCTGTGGGG GAAAAGCTAC
1201 TACAGGGAAG AGTGTCTCT GCTGTCTCTT CACTGGAAAA CAGGGAGGGG
1251 GGATTTGAGA CTGTGAAGAA AGTTGAATGG TGGTTTTTAA ATTATAAAGT
1301 AATGTATTAA AAGGTGCATT AGGCTGTAGT TCTAATATTG AGTTCAACTG
1351 TGAAATCCAT CAGATGTGCC AAATGGAGAA GACAGAAAGC AACAAAGTGA
1401 ATGTGTTCTT AGCCCAAGTG GTACAGTGAA TTTGCTTTAA CAGATGTTGA
1451 AAACATAAAT TTCTACTGTA TTCCCAGCAC GGGTGACTTC TTTTCTCTT
1501 CATTAGCCAG AGATGACTAA TTTAAATTTA GAACCAGATT TTAATTTAAA
1551 TTAATATTTT CATTAAATAA CTACTCATTT CAGATACCTA TTACTACTGTG
1601 TAACAGTTGT TTTGGAAATT TTATGTAAAA TTAACACTAT CAGTATTTTA
1651 CAGATGTTTT AATTAGACAT TGTATTAAAC AGGAACAGTG CAGAAACTAG
1701 AATCAAGCCT TATAATATCT TATAGACCAT GCATTTTGA AGTTAGTGTC
1751 CACTAGGGTC CTATTAAGTG TACATTGCA AGATTTTATT ATTTTGCCTT
1801 CTGACACTAT GGGAAAAAAT TTTTAGAAGC TATTGGGACA GATTCAAGCT
1851 TTTATGCACT TGGTTACTAC AGCTGTAAAA TGAAATCTCG TCTTGTAGCA
1901 TGGATTATTC TTCTCATGTT AAACCCACCA AAATAAAGGG GACTAAATAG
1951 GTAATGATTT TCCTAGTGCA TTTGCATACT GTGATAATCC TGGGCCTTGC
2001 AATAGTTCTA CAGGGCTCTT GGGCATTGAA TTATTAGGAT GTAATTGTAC
2051 ATCATTGTAG TGTTCACCTT ATTGAAGCTC ACTCTGATGT TAATGAGCTT
2101 CGGGTTTTGA TGCTTGTTTA GAGATCAGCA GTCTGGATG GGAGGGAACA
2151 AAGCTAAATA AATGTTAGTT TGGTGAAAAA AAAAAAAAAA AAAAA
```

BLAST Results

Entry HS599355 from database EMBL:
human STS WI-13484.

Score = 1262, P = 3.6e-52, identities = 274/289

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 901 bp; peptide length: 259
Category: similarity to unknown protein
Classification: no clue

```

1 MAACRALKAV LVDLSGTLHI EDAAVPGAQE ALKRLRGASV IIRFVTNTTK
51 ESKQDLLERL RKLEFDISED EIFTSLTAAR SLLERKQVRP MLLVDDRALP
101 DFKGIQTS DP NAVVMGLAPE HFHYQILNQA FRLLLDGAPL IAIHKARYYK
151 RKDGLALGPG PFVTALEYAT DTKATVVGKP EKTFFLEALR GTGCEPEEAV
201 MIGDDCRDDV GGAQDVGM LG ILVKTGKYRA SDEEKINPPP YLTCEFPFHA
251 VDHILQHLL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78d13, frame 2

TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid
K08B12., N = 1, Score = 609, P = 2.2e-59

TREMBL:CEC13C4_5 gene: "C13C4.4"; Caenorhabditis elegans cosmid C13C4,
N = 1, Score = 408, P = 4.4e-38

>TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid
K08B12.
Length = 257

HSPs:

Score = 609 (91.4 bits), Expect = 2.2e-59, P = 2.2e-59
Identities = 132/251 (52%), Positives = 172/251 (68%)

```

Query:      7 LKAVLVDSGLTHIEDAAVPGAQEALKRLRGASV IIRFVTNTTKESKQDLLERLRKLEFD 66
             + +VL+DLSGT+HIE+ A+PGAQ AL+ LR + + +FVTNTTKESK+ L +RL F
Sbjct:      4 ISSVLIDLSGTIHIEFAIPGAQTALRQHAQV-KFVTNTTKESKRLLHQRLINCGFK 62

Query:     67 ISEDEIFTSLTAARSLLERKQVRPMLLVDDRALPDFKGIQTS DPNAVVMGLAPEHFHYQI 126
             + ++EIFTSLTAAR L+ + Q RP +VDDRA+ DF+GI T DPNAVVMGLAPE F+
Sbjct:     63 VEKEEIFTSLTAARDLIVKNQYRPFIVDDRAMDFEGISTDDPNAVVMGLAPEKFNDDT 122

Query:    127 LNQAFLRLLDG-APLIAIHKARYYKRKDGALGPGPFVTALEYATDTKATVVGKPEKTF 185
             L AFRL+ + A LIAI+K RY++ GL LGPG +V LEY+ +AT+VGKP K FF
Sbjct:    123 LTHAFRLIKEKKASLIAINKGRYHQTNAGLC LGPGTYVAGLEYSAGVEATIVGKPNKLF 182

Query:    186 LEALRGTG--CEPEEAVMIGDDCRDDVGGAQDVGM LGILVKTGKYRASDEEKINPPPYLT 243
             AL+ + AVMIGDD DD GA +GM ILVKTGK+R DE K+
Sbjct:    183 ESALQSLNENVDFFSSAVMIGDDVNDALGAIKIGMRAILVKTGKFRDGLKVKN----V 238

Query:    244 CESFPHAVDHILQH 257
             SF AV+ I+++
Sbjct:    239 ANSFVDVNMIIEN 252

```

Pedant information for DKFZphfbr2_78d13, frame 2

Report for DKFZphfbr2_78d13.2

```

[LENGTH]      259
[MW]           28536.04
[pI]           5.84
[HOMOL]        TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid K08B12. 3e-
62
[FUNCAT]       r general function prediction [M. jannaschii, MJ1437] 3e-05
[SUPFAM]       nagD protein 4e-18
[KW]           Alpha_Beta

```

SEQ MAACRALKAVLVDSLGLHIEDAAVPGAQEALKRLRGASVIIRFVTNTTKESKQDLLERL
PRD cccccceeeeeecceeeccccchhhhhhhhhccceeeeeccccchhhhhhhh
SEQ RKLEFDISEDEIFTSLTAAARSLERKQVRPMLLVDDRALPDFKGIQTSDPNVVMGLAPE
PRD hhhccccceeeehhhhhhhhhhhccceeeeechhhhhccccccccceeecccc
SEQ HFHYQILNQAFRLLDGAPLIAIHKARYYKRKDGALGPGPFVTALEYATDTKATVVGKP
PRD chhhhhhhhhhhhhccceeeeeccccccccccccccccchhhhhhhccceeeccc
SEQ EKTFFLEALRGTCPEEAVMIGDDCRDDVGGAQDVGMGLILVKTGKYRASDEEKINPPP
PRD cchhhhhhhhhhhccceeeeeccccchhhhhhhhhccceeeeecccccccccccc
SEQ YLTCEFPFAVDHILQHLL
PRD cccccchhhhhhhhhcc

(No Prosite data available for DKFZphfbr2_78d13.2)

(No Pfam data available for DKFZphfbr2_78d13.2)

DKFZphfbr2_78k24

group: metabolism

DKFZphfbr2_78k24 encodes a novel 372 amino acid protein with similarity to Mus musculus ubiquitin specific protease UBP43.

The novel protein contains a Prosite ubiquitin carboxyl-terminal hydrolases family 2 signature 2. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins.

The new protein can find application in modulation of protein stability/degradation in cells.

Ubiquitin carboxyl-terminal hydrolases family 2 signature 2.

strong similarity to mouse ubiquitin specific protease UBP43

Sequenced by MediGenomix

Locus: unknown

Insert length: 1874 bp

Poly A stretch at pos. 1852, polyadenylation signal at pos. 1836

```
1 AGTCCCGACG TGGAACTCAG CAGCGGAGGC TGGACGCTTG CATGGCGCTT
51 GAGAGATTCC ATCGTGCCTG GCTCACATAA GCGCTTCCTG GAAGTGAAGT
101 CGTGTGTGCC TGAACGCGGG CCAGGCAGCT GCGGCTTGGG GGTTTTGGAG
151 TGATCACGAA TGAGCAAGGC GTTTGGGCTC CTGAGGCAAA TCTGTCAGTC
201 CATCCTGGCT GAGTCCTCGC AGTCCCGGCG AGATCTTGAA GAAAAGAAGG
251 AAGAAGACAG CAACATGAAG AGAGAGCAGC CCAGAGAGCG TCCCAGGGCC
301 TGGGACTACC CTCATGGCCT GGTGGTTTA CACAACATTG GACAGACCTG
351 CTGCCCTTAA TCCTTGATTG AGGTGTTTCT AATGAATGTG GACTTCACCA
401 GGATATTGAA GAGGATCAGC GTGCCCAGGG GAGCTGACGA GCAGAGGAGA
451 AGCGTCCCTT TCCAGATGCT TCTGTGCTG GAGAAGATGC AGGACAGCCG
501 GCAGAAAGCA GTGCGGCCCC TGGAGCTGGC CTACTGCCTG CAGAAAGTGA
551 ACGTGGCCCT GTTTGTCCAA CATGATGCTG CCCAACTGTA CCTCAAACCTC
601 TGGAACTTGA TTAAGGACCA GATCACTGAT GTGCACTTGG TGGAGAGACT
651 GCAGGCCCCG TATACGATCC GGGTGAAGGA CTCCTTGATT TGCCTTGACT
701 GTGCCATGGA GAGTAGCAGA AACAGCAGCA TGCTACCCCT CCCACTTTCT
751 CTTTTTGATG TGGACTCAAA GCCCCTGAAG AACTGGAGG ACGCCCTGCA
801 CTGCTTCTTC CAGCCCAGGG AGTTATCAAG CAAAAGCAAG TGCTTCTGTG
851 AGAAGTGTGG GAAGAAGACC CGTGGGAAAC AGGTCTTGAA GCTGACCCAT
901 TTGCCCCAGA CCTGACAAAT CCACCTCATG CGATTCTCCA TCAGGAATTC
951 ACAGACGAGA AAGATCTGCC ACTCCCTGTA CTCCCCCAG AGCTTGGATT
1001 TCAGCCAGAT CCTTCCAATG AAGCGAGAGT CTGTGATGTC TGAGGAGCAG
1051 TCTGGAGGGC AGTATGAGCT TTTTGCTGTG ATTGCGCAGC TGGGAATGGC
1101 AGACTCCGGT CATTACTGTG TCTACATCCG GAATGCTGTG GATGGAAAAT
1151 GGTTCGTGCT CAATGACTCC AATATTGCTT TGGTGTCTCT GGAAGACATC
1201 CAGTGTACCT ACGGAAATCC TAACTACCAC TGGCAGGAAA CTGCATATCT
1251 TCTGGTTTAC ATGAAGATGG AGTGCTAATG GAAATGCCCA AAACCTTCAG
1301 AGATTGACAC GCTGTCATTT TCCATTTCCT TTCCTGGATC TACGGAGTCT
1351 TCTAAGAGAT TTTGCAATGA GGAGAAGCAT TGTTTTCAAA CTATATAACT
1401 GAGCCTTATT TATAATTAGG GATATTATCA AAATATGTAA CCATGAGGCC
1451 CCTCAGGTCC TGATCAGTCA GAATGGATGC TTTCACCAGC AGACCCGGCC
1501 ATGTGGCTGC TCGGTCTGCG GTGCTCGCTG CTGTGCAAGA CATTAGCCCT
1551 TTAGTTATGA GCCTGTGGGA ACTTCAGGGG TTCCAGTGGG GGAGAGCAGT
1601 GGCAGTGGGA GGCATCTGGG GGCCAAAGGT CAGTGGCAGG GGGTATTTC
1651 GTATTATACA ACTGCTGTGA CCAGACTTGT ATACTGGCTG AATATCAGTG
1701 CTGTTTGTAA TTTTCACTT TGAGAACCAA CATTAAATCC ATATGAATCA
1751 AGTGTTTTGT AACTGCTATT CATTATTTCA GCAAATATTT ATTGATCATC
1801 TCTTCTCCAT AAGATAGTGT GATAAACACA GTCATGAATA AAGTTATTTT
1851 CCACAAAAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry AC005500 from database EMBL:

, complete sequence.

Score = 859, P = 5.7e-143, identities = 175/179

8 exons matching Bp 317-1230

Medline entries

99182491:

A novel ubiquitin-specific protease, UBP43, cloned from leukemia fusion protein AML1-ETO-expressing mice, functions in hematopoietic cell differentiation.

Peptide information for frame 1

ORF from 160 bp to 1275 bp; peptide length: 372
 Category: strong similarity to known protein
 Classification: Protein management
 Prosite motifs: UCH_2_2 (302-320)

```

1 MSKAFGLLRQ ICQSILAESS QSPADLEEKK EEDSNMKREQ PRERPRAWDY
51 PHGLVGLHNI GQTCCNLNLI QVFVMNVDFTRILKRITVPR GADEQRRSVP
101 FQMLLLLEKM QDSRQKAVRP LELAYCLQKC NVPLFVQHDA AQLYLKLWNL
151 IKDQITDVHL VERLQALYTI RVKDSLICVD CAMESSRNSS MLTLPPLSLFD
201 VDSKPLKTL DALHCFQPR ELSSKSKCFC ENCGKKTRGK QVLKLTHLPQ
251 TLTIHLMRFS IRNSQTRKIC HSLYFPQSLD FSQILPMKRE SCDAEEQSGG
301 QYELFAVIAH VGMADSGHYC VYIRNAVDGK WFCFNDNSNIC LVSWEIDIQT
351 YGNPNYHWQE TAYLLVYMK EC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78k24, frame 1

TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus ubiquitin specific protease UBP43 mRNA, complete cds., N = 1, Score = 1367, P = 1e-139

SWISSPROT:UBPE_DROME UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 64E) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 64E) (DEUBIQUITINATING ENZYME 64E)., N = 2, Score = 248, P = 5.3e-33

>TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus ubiquitin specific protease UBP43 mRNA, complete cds.
 Length = 368

HSPs:

Score = 1367 (205.1 bits), Expect = 1.0e-139, P = 1.0e-139
 Identities = 262/369 (71%), Positives = 295/369 (79%)

```

Query:      1 MSKAFGLLRQICQSILAESSQSPADLEEKK EEDSNMKREQPRERPRAWDYPHGLVGLHNI 60
            M K FGLLR+ CQS++AE Q A LEE E   KR R+ AWD PHGLVGLHNI
Sbjct:      1 MGKGFGLLRKPQSVVAEPQQYSA-LEE--ERTMKRKRVLSDLCASAWDSPHGLVGLHNI 57

Query:      61 GQTCCNLNLIQVFVMNVDFTRILKRITVPRGADEQRRSVPFQMLLLLEKMQDSRQKAVRP 120
            GQTCCNLNL+QVF+MN+DF ILKRITVPR A+E++RSVPFQ+LLLLLEKMQDSRQKA+ P
Sbjct:      58 GQTCCNLNLLQVFMMNMDFRMILKRITVPRSAEERKRSVPFQLLLLLEKMQDSRQKALLP 117

Query:      121 LELAYCLQKCNVPLFVQHDAQAQLYLKLWNLIKDQITDVHLVERLQALYTIIRVDSLICVD 180
            EL CLQK NVPLFVQHDAQAQLYL +WNL KDQITD L ERLQ L+TI ++SLICV
Sbjct:      118 TELVQCLQKYNVPLFVQHDAQAQLYLTINWNLTKDQITDLDLTERLQGLFTIWTQESLICVG 177

Query:      181 CAMESSRNSSMLTLPPLSLFDVDSKPLKTLLEDALHCFQPRELSSKSKCFCENCGKKTRGK 240
            C ESSR S +LTL L LFD D+KPLKTLLEDAL CF QP+EL+S   C CE CG+KT K
Sbjct:      178 CTAESSRRSKLLTSLPLPLFDKDAKPLKTLLEDALRCFVQPKELASSDMC-CETCGEKTTPWK 236

Query:      241 QVLKLTHLPQTTLTIHLMRFSIRNSQTRKICHSLYFPQSLDFSQILPMKRESCDAEEQSGG 300
            QVLKLTHLPQTTLTIHLMRFS RNS+T KICHS+ FPQSLDFSQ+LP + + D +EQS
Sbjct:      237 QVLKLTHLPQTTLTIHLMRFSARNRSRTEKICHSVNFQSLDFSQVLPTEEDLGDTKEQSEI 296

Query:      301 QYELFAVIAHVGMAADSGHYCVYIRNAVDGKWFCFNDNSNICLVSWEIDIQTYGNPNYHWQE 360
            YELFAVIAHVGMAAD GHYC YIRN VDGKWFCFNDNS++C V+W+D+QCTYGN Y W+E
Sbjct:      297 HYELFAVIAHVGMAADFGHYCAYIRNPVDGKWFCFNDNSHVCWVTWKDVQCTYGNHRYRWRE 356

Query:      361 TAYLLVYMK 369

```



```
HMM_NAME      Ubiquitin carboxyl-terminal hydrolases family 2
HMM            *YdLYgVICHYGntldyGHYwaYVKNenhHRWkWYYFDDEtv*
               Y+L++VI H G   D+GHY +Y++N   ++KW++F+D+++
Query          302 YELFAVIAHVG-MADSGHYCVYIRNAV--DGKWFCFNDSDNI 339
```

DKFZphfbr2_78n23

group: brain derived

DKFZphfbr2_78n23 encodes a novel 329 amino acid protein with similarity to A.thaliana F26P21.80 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F26P21.80

Sequenced by MediGenomix

Locus: /map="89.1 cR from top of Chr19 linkage group"

Insert length: 1447 bp

Poly A stretch at pos. 1374, polyadenylation signal at pos. 1353

```
1 TACAACCTCC GGCTGTAAAG ATGGCGGCTT CCTAGTGAGT CGGCGGCTGA
51 CTTAGAAGGA GGTTCAGGCT ACGGTGAGCC GAAGCCACAC AGGAGCCATG
101 GAAGTGGCAG AGCCCAGCAG CCCCACTGAA GAGGAGGAGG AGGAAGAGGA
151 GCACTCGGCA GAGCCTCGGC CCCGCACTCG CTCCAATCCT GAAGGGGCTG
201 AGGACCGGGC AGTAGGGGCA CAGGCCAGCG TGGGCAGCCG CAGCGAGGGT
251 GAGGGTGAGG CCGCCAGTGC TGATGATGGG AGCCTCAACA CTTCAGGAGC
301 CGGCCCTAAG TCCTGGCAGG TGCCCCCGCC AGCCCCTGAG GTCCAAATTC
351 GGACACCAAG GGTCAACTGT CCAGAGAAAG TGATTATCTG CCTGGACCTG
401 TCAGAGGAAA TGTCACTGCC AAAGCTGGAG TCGTTCACAG GCTCCAAAAC
451 CAACGCCCTC AATGTCTCTC AGAAGATGAT TGAGATGTTT GTGCGGACAA
501 AACACAAGAT CGACAAAAGC CACGAGTTTG CACTGGTGGT GGTGAACGAT
551 GACACGGCCT GGCTGTCTGG CCTGACCTCC GACCCCGCGG AGCTCTGTAG
601 CTGCCTCTAT GATCTGGAGA CGGCCTCCTG TTCCACCTTC AATCTGGAAG
651 GACTTTTCAG CCTCATCCAG CAGAAAAGTG AGCTTCGGGT CACAGAGAAC
701 GTGCAGACGA TTCCCCCGCC ATATGTGGTC CGCACCATCC TTGTCTACAG
751 CCGTCCACCT TGCCAGCCCC AGTTCTCCTT GACGGAGCCC ATGAAGAAAA
801 TGTTCACAGT CCCATATTTC TTCTTTGACG TTGTTTACAT CCACAATGGC
851 ACTGAGGAGA AGGAGGAGGA GATGAGTTGG AAGGATATGT TTGCCTTCAT
901 GGGCAGCCTG GATACCAAGG GTACCAGCTA CAAGTATGAG GTGGCACTGG
951 CTGGGCCAGC CTTGGAGTTG CACAACCTGA TGGCGAAACT GTTGGCCAC
1001 CCCCTGCAGC GGCCTTGCCA GAGCCATGCT TCCTACAGCC TGCTGGAGGA
1051 GGAGGATGAA GCCATTGAGG TTGAGGCCAC TGTCTGAACC ATCCCTGTAC
1101 ATCTGCACCT TCTTGTGCAA GGAAGTCCTT GGCCTAAAGC CTGGGTTCTC
1151 AAAGTGGGTT CTTGGGACCC TCCGGGGTGG GGGGGTTCCA GGAGGCACGT
1201 AGGGTACCTT GCAGGGTCTT AGGAGGGAAA CCCAGGATTC CAGGAGGGAT
1251 CCCAGGAAGT GTGGGCACCC ATTTTCTGTG TCTCCAGGCC CATTTCCACT
1301 CCTAGTTTGT CATGGATAAT TTTTGTCTT CCCTGTGTGA TTTTGGCCAT
1351 CAAAATAAAA ATTTGAGACT CGTTAAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAAGAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry HS806352 from database EMBL:
human STS EST192543.
Score = 1285, P = 2.5e-51, identities = 263/266

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 98 bp to 1084 bp; peptide length: 329
Category: similarity to unknown protein
Classification: no clue

1 MEVAEPSSPT EEEEEEEHS AEPRPRTRSN PEGAEDRAVG AQASVGSRSR

```

51 GEGEASADD GSLNTSGAGP KSWQVPPAP EVQIRTPRVN CPEKVIICLD
101 LSEEMSLPKL ESFNGSKTNA LNVSKMIEM FVRTKHKIDK SHEFALVVVN
151 DDTAWLSGLT SDPRELCSCL YDLETASCST FNLEGLFSLI QKQTELPVTE
201 NVQTIPPPYV VRTILVYSRP PCQPQFSLTE PMKKMFQCPY FFFDVVYIHN
251 GTEEKEEEMS WKDMFAFMGS LDTKGTSYKY EVALAGPALE LHNMAKLLA
301 HPLQRPCQSH ASYSLLEED EAIEVEATV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78n23, frame 2

PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana, N = 1, Score = 142, P = 1.5e-07

>PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana
Length = 264

HSPs:

Score = 142 (21.3 bits), Expect = 1.5e-07, P = 1.5e-07
Identities = 56/216 (25%), Positives = 97/216 (44%)

```

Query:   93 EKVIICLDL-SEEMSLPKLESFNGSKTNALNVSKMIEMFVRTKHKIDKSHEFALVVND 151
          E ++IC+D+ +E M K NG + ++ I +F+ K I+ H FA +
Sbjct:   26 EDILICIDVDAESMVEMKTTGTNGRPLIRMECVKQAIILFIHNKLSINPDHRFAFATLAK 85

Query:   152 DTAWLSG-LTSDPRELCSCLYDLE-TASCSTFNLEGLFSLIQKQTELPVTENVQTIPPPY 209
          AWL TSD + L L S S +L LF Q+ ++ +N
Sbjct:   86 SAAWLKKEFTSDAESAVASLRGLSGNKSSSRADLTLLFRAAAQEAQVSRQN-----R 138

Query:   210 VVRTILVYSRPPCQPQFSLTEPMKKMFQCPYFFFDVVYIHNNGTEEKEEEMSWKDMF-AFM 268
          + R IL+Y R +P P+ + F DV+Y+H ++ + +D++ + +
Sbjct:   139 IFRVILIYCRSSMRPTHEW--PLNQKL---FTLDVMYLH---DKPSPDNCPPQDVYDSL 189

Query:   269 GSLD--TKGTSYKYEVALAGPALELHNMAKLLAHPLQRPCQ 308
          +++ ++ Y +E G A + M+ LL HP QR Q
Sbjct:   190 DAVEHVSEYEGYIFESG-QGLARSVFKPMSMLLTHPQORCAQ 230

```

Pedant information for DKFZphfbr2_78n23, frame 2

Report for DKFZphfbr2_78n23.2

```

[LENGTH] 329
[MW]      36560.10
[pI]      4.60
[HOMOL]   PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana 7e-07
[KW]      Alpha_Beta
[KW]      LOW_COMPLEXITY 9.73 %

```

```

SEQ MEVAEPSSPTEEEEEEHSAEPRPRTSRNPEGAEDRAVGAQASVGSRSEGEAGEASADD
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhhcccccccccccccc

SEQ GSLNTSGAGPKSWQVPPAPAEVQIRTPRVNCPEKVIICLDLSEEMSLPKLESFNGSKTNA
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LNVSKMIEMFVRTKHKIDKSHEFALVVNDTAWLSGLTSDPRELCSCLYDLETASCST
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ehhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhcccccccccc

SEQ FNLEGLFSLIQKQTELPVTENVQTIPPPYVVRTILVYSRPPCQPQFSLTEPMKKMFQCPY
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD hhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhheeee

SEQ FFFDVVYIHNNGTEEKEEEMSWKDMFAFMGSLDTKGTSYKYEVALAGPALELHNMAKLLA
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD eeeeeeeccccchhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhh

SEQ HPLQRPCQSHASYSLLEEDAEIEVEATV
SEG .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD hccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

(No Prosite data available for DKFZphfbr2_78n23.2)

(No Pfam data available for DKFZphfbr2_78n23.2)

DKFZphfbr2_7a24

group: brain derived

DKFZphfbr2_7a24 encodes a novel 142 amino acid protein with similarity to the C-terminal part of transforming growth factor-beta activated kinases.

The novel protein shows only similarity to the C-terminus of such kinases; no kinase domain is present.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C-terminus of TGF-beta-activated kinase

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1697 bp

No poly A stretch found, no polyadenylation signal found

```
1 GGGGAGAGAG GGGTTGTGAA GGAAGCGGA AGGAAGGGA AGGGAGGTCC
51 CGTGGGACGC TGGGGTCTGG GGTAGAGCAG GTAGCAGCGT GCTGCCCTGA
101 CAGCTGTCTC CGCTCCTCAG ATTGTCAAGT GCTGCTATGC AGCAGGTGCA
151 GCCTGGTCTC TCACTGAGTC TCTACTCCAC AAAGGCAACG ACTGGCCAAG
201 GCAGTGGCTG GCTCTGGGTT ACACAAGTGC AGACACTCAA CTAAGTGAGC
251 TGGAAGACCC AGGAGAAGGC GGAGGCTCAG GTGCCACAT GATCAGCACA
301 GCCAGGGTAC CTGCTGACAA GCCTGTACGC ATCGCCTTTA GCCTCAATGA
351 CGCCTCAGAT GATACACCCC CTGAAGACTC CATTCTTTTG GTCTTTCCAG
401 AATTAGACCA GCAGCTACAG CCCCTGCCGC CTTGTCATGA CTCCGAGGAA
451 TCCATGGAGG TGTTGAGACA GCACTGCCAA ATAGCAGAAG AATACCTTGA
501 GGTCAAAAAG GAAATCACCC TGCTTGAGCA AAGGAAGAAG GAGCTCATTG
551 CCAAGTTAGA TCAGGCAGAA GAGGAGAAGG TGGATGCTGC TGAGCTGGTT
601 CGGGAATTGC AGGCTCTGAC GGAGGAGAAT CGGACGTTGA GGTGGGCCCA
651 GTCTCAATGT GTGGAACAAC TGGAGAAACT TCGAATACAG TATCAGAAGA
701 GGCAGGGCTC GTCCTAACTT TAAATTTTTC AGTGTGAGCA TACGAGGCTG
751 ATGACTGCCC TGTGCTGGCC AAAAGATTTT TATTTTAAAT GAATAGTGAG
801 TCAGATCTAT TGCTTCTCTG TATTACCCAC ATGACAACATG TCTATAATGA
851 GTTTACTGCT TGCCAGCTTC TAGCTTGAGA GAAGGGATAT TTTAATGAG
901 ATCATTAAAC TGAAGTATT ACTAGTATAT GTTTTGGAG ATCAGAATTC
951 TTTTCCAAAG ATATATGTTT TTTTCTTTT TAGGAAGATA TGATCATGCT
1001 GTACAACAGG GTAGAAAATG GTAAAATAG ACTATTGACT GACCCAGCTA
1051 AGAATCGCGG GCTGAGCAGA GTTAAACCAT GGGACAAACC CATAACATGT
1101 TCACCATAGT TTCACGTATG TGTATTTTAA AATTTTCATGC CTTTAATATT
1151 TCAAAATATG TCAAAATTTAA ACTGTCAGAA ACTTCTCTGC ATGTATTTAT
1201 ATTTGCCAGA GTATAAACTT TTATACTCTG ATTTTATCC TTCAATGATT
1251 GATTATACTA AGAATAAATG GTCACATATC CTAAGAGCTT CTTTATGAAA
1301 TTATTAGCAG AAACCATGTT TGAAACCAAA GCACATTTGC CAATGCTAAC
1351 TGGCTGTTGT AATAATAAAC AGATAAGGCT GCATTGCTT CATGCCATGT
1401 GACCTCACAG TAAACATCTC TGCCTTTGCC TGTGTGTGTT CTGGGGGAGG
1451 GGGGACATGG AAAAATATTG TTTGGACATT ACTTGGGTGA GTGCCCATGA
1501 AGACATCAGT GAACTTGTA CTATTGTTT GTTTTGGATT TAAGGAGATG
1551 TTTTAGATCA GTAACAGCTA ATAGGAATAT GCGAGTAAAT TCAGAATTGA
1601 AACAATTTCT CTTGTTCTA CTTATCACCA CATTCTTCTA AATTGAATCT
1651 TTTGTATAT GTCCATTCT ATTCTGTAA CTCTTTTTC ATTAAC
```

BLAST Results

No BLAST result

Medline entries

98130593:
Role of TAK1 and TAB1 in BMP signaling in early Xenopus development.

Peptide information for frame 1

ORF from 289 bp to 714 bp; peptide length: 142
Category: similarity to known protein

1 MISTARVPAD KPVRIAFSLN DASDDTPPED SIPLVFPELD QQLQPLPPCH
51 DSEESMEVFR QHCQIAEEYL EVKKEITLLE QRKKELIAKL DQAEEEKVDA
101 AELVREFEAL TEENRTLRLA QSQCVEQLEK LRIQYQKRQG SS

BLASTP hits

Entry U92030_1 from database TREMBL:
product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1 mRNA,
complete cds.
Score = 343, P = 1.3e-30, identities = 69/143, positives = 104/143

Entry AB009356_1 from database TREMBL:
product: "TGF-beta activated kinase 1a"; Homo sapiens mRNA for
TGF-beta activated kinase 1a, complete cds.
Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry MMPK_1 from database TREMBL:
product: "TAK1 (TGF-beta-activated kinase)"; Mouse mRNA for TAK1
(TGF-beta-activated kinase), complete cds.
Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry AB009357_1 from database TREMBL:
product: "TGF-beta activated kinase 1b"; Homo sapiens mRNA for
TGF-beta activated kinase 1b, complete cds.
Score = 339, P = 3.2e-30, identities = 67/143, positives = 104/143

Entry AB009358_1 from database TREMBL:
product: "TGF-beta activated kinase 1c"; Homo sapiens mRNA for
TGF-beta activated kinase 1c, complete cds.
Score = 144, P = 3.8e-09, identities = 30/67, positives = 47/67

Alert BLASTP hits for DKFZphfbr2_7a24, frame 1

PIR:JC5955 transforming growth factor-beta activated kinase (EC
-.-.-.-) 1a - Human, N = 1, Score = 339, P = 3e-30

>PIR:JC5955 transforming growth factor-beta activated kinase (EC -.-.-.-) 1a
- Human

Length = 579

HSPs:

Score = 339 (50.9 bits), Expect = 3.0e-30, P = 3.0e-30
Identities = 67/143 (46%), Positives = 104/143 (72%)

Query: 1 MISTARVPADKPVRI-AFSLNDASDDTPPEDSIPLVFPELDQQLQPLPPCHDSEESMEVF 59
MI+T+ ++KP R ++ +D++D ++SIP+ + LD QLQPL PC +S+ESM VF
Sbjct: 437 MITTS GPTSEKPTRSHPTPDDSTDTNGSDNSIPMAYLTLDHQLQPLAPCPNSKESMAVF 496

Query: 60 RHCQIAEEYLEVKKEITLLEQRKKELIAKLDQAEEEKVDAAELVREFEALTEENRTLRL 119
QHC++A+EY++V+ EI LL QRK+EL+A+LDQ E+++ + + LV+E + L +EN++L
Sbjct: 497 EQHCKMAQEYMKVQTEIALLLQRKQELVAELDQDEKQQNTSRLVQEHHKLLDENKSLST 556

Query: 120 AQSQCVEQLEKLRIQYQKRQGSS 142

Sbjct: 557 YYQCKKQLEVIRSQQKQROGTS 579

Pedant information for DKFZphfbr2_7a24, frame 1

Report for DKFZphfbr2_7a24.1

[LENGTH] 142
[MW] 16377.53
[pI] 4.64
[HOMOL] TREMBL:U92030_1 product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1
mRNA, complete cds. 6e-26
[PROSITE] CK2_PHOSPHO_SITE 3

```

[PROSITE]      PKC_PHOSPHO_SITE      2
[PROSITE]      ASN_GLYCOSYLATION     1
[PFAM]         TNFR/NGFR cysteine-rich region
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY         7.04 %
[KW]           COILED_COIL           33.10 %

```

```

SEQ      MISTARVPADKPVRIAFSLNDASDDTPPEDSIPLVFPELDQQLQPLPPCHDSEESMEVFR
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....

```

```

SEQ      QHCQIAEEYLEVKKEITLLEQRKKELIAKLDQAEEKVDAAELVREFEALTEENRTLRLA
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    ...cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      QSQCQEQLKLRIQYQKRQGSS
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....

```

Prosites for DKFZphfbr2_7a24.1

PS00001	114->118	ASN_GLYCOSYLATION	PDOC00001
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00006	18->22	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006

Pfam for DKFZphfbr2_7a24.1

```

HMM_NAME      TNFR/NGFR cysteine-rich region
HMM            *CpeGtYtDWNHvpqClpCtrCePEMGQYMvqPCTwTQNTVC*
               C++++ + + +Q C++ E+ +++++ T + ++
Query          49 CHDSEESMEVF-RQH--CQIAEE--YLEVKKEITLLEQRKK      84

```

DKFZphfbr2_7e22

group: brain derived

DKFZphfbr2_7e22.2 encodes a novel 286 amino acid protein similar to b561 cytochromes

The new protein shows strong similarity to B561 cytochromes, but contains no heme binding site. In addition, a myc-type, helix-loop-helix dimerization domain domain is present. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to cytochrome b561

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 4254 bp

Poly A stretch at pos. 4234, polyadenylation signal at pos. 4217

```

1  GGGGACTACC CAGAGGGCTG CCGCGGCTC TCCAAGTTCT TGTGGCCCCC
51  GCGGTGCGGA GTATGGGGCG CTGATGGCCA TGGAGGGCTA CCGGCGCTTC
101 CTGGCGCTGC TGGGGTCGGC ACTGCTCGTC GGGTTCCTGT CCGTGATCTT
151 CGCCCTCGTC TGGGTCTCTC ACTACCGAGA GGGGCTTGGC TGGGATGGGA
201 GCGCACTAGA GTTAACTGG CACCCAGTGC TCATGGTTCAC CGGCTTCGTC
251 TTCCATCCAGG GCATCGCCAT CATCGTCTAC AGACTGCCGT GGACCTGGAA
301 ATGCAGCAAG CTCCTGATGA AATCCATCCA TGCAGGGTTA AATGCAGTTG
351 CTGCCATTCT TGCAATTATC TCTGTGGTGG CCGTGTTTGA GAACCACAAT
401 GTTAAACAATA TAGCCAAATAT GTACAGTCTG CACAGCTGGG TTGGACTGAT
451 AGCTGTGATA TGCTATTGTG TACAGCTTCT TTCAGGTTTT TCAGTCTTTC
501 TGCTTCCATG GGCTCCGCTT TCTCTCCGAG CATTTCTCAT GCCCATACAT
551 GTTTATTCTG GAATTGTCAT CTTTGAACA GTGATTGCAA CAGCACTTAT
601 GGGATTGACA GAGAACTGA TTTTTCCTT GAGAGATCCT GCATACAGTA
651 CATTCGCCGC AGAAGGTGTT TTCGTAATA CGCTTGGCCT TCTGATCCTG
701 GTGTTGCGGG CCTCATTTT TTGGATAGTC ACCAGACCGC AATGGAAACG
751 TCCTAAGGAG CCAAATCTA CCATTCTTCA TCCAAATGGA GGCACCTGAA
801 AGGGAGCAAG AGGTTCCATG CCAGCCTACT CTGGCAACAA CATGGACAAA
851 TCAGATTGAG AGTTAAACAA TGAAGTAGCA GCAAGGAAAA GAAACTTAGC
901 TCTGGATGAG GCTGGGCAGA GATCTACCAT GTAAATGTT GTAGAGATAG
951 AGCCATATAA CGTCACGTTT CAAAACCTAGC TCTACAGTTT TGCTTCTCCT
1001 ATTAGCCATA TGATAATTGG GCTATGTAGT ATCAATATTT ACTTTAATCA
1051 CAAAGGATGG TTTCTGAAA TAATTGTATG TGATTGAGGC CTATGAACTG
1101 ACCTGAATTG GAAAGGATGT GATTAATATA AATAATAGCA GATATAAATT
1151 GTGGTTATGT TACCTTTATC TTGTTGAGGA CCACAACATT AGCACGGTGC
1201 CTTGTGCAGA ATAGATACTC AATATGTGAA TATGTGTCTA CTAGTAGTTA
1251 ATTGGATAAA CTGGCAGCAT CCCTGGCCTG TTGTCATGCA GTCATTTCTT
1301 GTTAATTCTG GGAGACAATG ATTTCAACAC TAGAGGGAAG CAGTCTTAAA
1351 AGTTTAAAT CCGATAAGGA ATATCTGGGA CAGGGTTTAG ATCATGACTC
1401 TACACAGATA CCATGATGAG AGTATATTAA AGAAATTTAG GAAAGCACCT
1451 GGTTCCTTTT TCCCCATGCC TGCCTTCTGC TCCCTCCCCA GCTGGTTTGG
1501 GCTCAAATTG TCCCTGGAGA CTAGGGTTTA TGTTAGGGTA TTGATAGATT
1551 AGAGCAGGTG GTTGAAGAGA TCTTCTCTGG TCAGACTTGG AAGAATTTC
1601 AAAAGTGAAG TTAGCCCCAA GACTTCCCTA GGGTTGATGT ACTTTATGAT
1651 CCAGATGCTA AACTTCTTAG AATGAAAATA TGCTTCAACA CTTAAGTAGC
1701 ATACACTGCC CTACAAACCT CAGAGAGCAC TTTTCCCCAA GTTCTTGTTT
1751 TTATTTTGA AAGTACTCAC ACAGCACTTA CTATGCTCCA AACACTCCTC
1801 TAAGCACTTT ACACATATTA GCTCATTTCAG TCCCCAGACA GACGGGATGA
1851 AGTAGGTATT GTTACTGTTT CCATTTTACA GGTGAGAGAT TTGAAGCCTG
1901 GGGAGGCTAG TAACTACCC CAAGGTCACA CGGCTCATAC ATGGTGGGAC
1951 TGAGACTCAG ATGCAGGCAG TCTGGCACCT CAGTCTGGAT TCTAACCAT
2001 TCACTAAGCT ATTTTGTCT TGTACTACTT TGACCCACCC CTGAATAAAC
2051 CTCATTTGCT GGAGTGGGGT GTAGTTATTA AAGGGATGCT TTTTACCTTT
2101 TGCTGTCTGC TGTGGCAGAT TCCCCAGATA ACCAAGGAAA AGGGGCCACC
2151 CATACCTGGA AATAGGCCAT AGGGCCCCTA CTACTGCCAA CAAGCCATGG
2201 CCTACCTTGA CACTTGTGTT ATCTTAAAT TGTGTCTTGG TAACAAAAGA
2251 TTTGGACAGG CATATCTGTA GCTTTCAAGT TAATTAATTG CAATATTTT
2301 TTCTTCAGGA TTTTAGCTGC TGAACAACCT TCAGTTTGA GCTAAAAGAG
2351 ACCTGTCTCA TGGTCTGCCC TTCCCTGGGG CAATAGCTAG GGTCTTTCCT
2401 GATTTTTATG GAATTTTAGG GGATATTTTG AGCTTTGGGT TCTCAGTAGT

```

```

2451 GAATTGAGAC TTGGAGGTGA CTTTTCATGT TTGGAGTATC ATCTCTGTCT
2501 GGGCTCTGGG CTGACAAATT AAAACCTAGA GTAGTGCTTA TGCTGAAATG
2551 ATACTTTTCA TTTTGTGGT GATTTTTTTG CCTTCCCTTC AATTTTAAAC
2601 TGAAGCATT TAATGTGGGT AGAAACTCTA CACCAAAATC ACTAAACATT
2651 TTGGTGCTTA GTGGATTCT TTTTAGGTAA CTGGTACTTA CTTCCAAAGA
2701 CTGAATACAA GCCACACTCC ATCATATCCC TTAACCTCA TGAAAAACCA
2751 TTCAAGATCC CCTTGCTGCA AACTGTCTCT CTCTCTCTCT ACTAAATCT
2801 ATTTCCAAAA TTGGTAATAG AGCCAGAAGG ATCCCAGTA CCCAGCCCTC
2851 TGCCTGGCAC AAAGTGGTAG CACAATTAAA TTCAGTATGG GTGGAGCATG
2901 GTACAGTCTT GGTGCCATAG AAGGAGTAGT TGCATAGTCA CACATCATTT
2951 GATAAGTTGG ATGTTCATT ACATAGAGGA ACACAAATTT CCAGGGTTTT
3001 TGGAGGAAGG GATTAGTAG CGACTAAGCC GCCAGAATTG AGGTGGCCAT
3051 TCCTTTTGT ATAGGCTAAG AAACAGGTTA TCAGTGAAAA GTTAATTATG
3101 GCTTTGGCAC TAGAATAGCA CTGTGCAAA GTATTTAAGC ACCCCCCATC
3151 TCAGCCCTTT ATTTATCTT TCATGTGGG TAATGTGAGG ATAATCTTAC
3201 AGATATTATA GGAATTTCTT TTCTATCTT ATGAAACAA CGTATATAAA
3251 ATATATCTAG AAAACCTTTG TTTGAGACTC TTATTTAATG GGCTTTTGAT
3301 TCTAATGATA ATTGTACCTT TATCTTTCAA AAGCTGATAT TTCTACCTA
3351 AGCATCTCCC GAGAAAAATA TCTCATTAAA AAGCCCATAA ATAATAGGGG
3401 AGAAGAAAGC CTTAGGTATC AATCCAAAA CAGTGATTGA AATTTCCCAA
3451 AATAATTATG GCTTCTGTCA TCTCCAGAGA TAATCTGGCT TGGTTTACCC
3501 CATAATCTAA TTTCCAGAAA GAAAGCTTTA TTTTAACACT CATCTGAATC
3551 AACATTAAAG CCTTTCTCT CAAAGCGTTT ATTGAGAAAC TCAAATGAAT
3601 ATACTTTTTG AATTACTGTC ATCAAAAGTG TACGGCTTCC TGTGCTGCTT
3651 GTGTCAAATG GAACCTGCCC TCTAAAGCAC TTTCTTTCCCT TTACTTGCGT
3701 GGTTCATGT AAGCTGTGCT GTTTAGAAAC AACATCTCAG ACTTTACAAA
3751 GAAATGACAA AGAAGGCAAT TGCACTTTTT AAGGGATATC GACAAGCAGT
3801 TTTCTGTTTC TAAAGGACAA AATACAGAGT GTGTGTCAAT TTTAATTAGA
3851 TTTTCCCCC TGCTGAGTTG GAAATCCAG TGCAGCACTG ATTGACCACA
3901 GTTGCCAATC TAAAGCACAA AAGACAGAAG TAAAGCTTTA TGCTAATTTT
3951 ATTTCAATAT GATAGAAAAT TTATCTTGGT ATGTCTTTT TTAGATAACT
4001 CCAGCAGGAA ACTGTAACTG CTATGTCTTT AGGAAAACGT AGAAGAAAGA
4051 ACATTATTAT TCTTTAATTC CTACAAGGTA CTTGAAACC TTAAGTGAAA
4101 AAGATTCTTA TCTTTTATC TTGGCGCATT TATGAAAAA ATATTAACTG
4151 TCCTGAATAT TTTATAATTT TGTAGGAAAA ATATGCATCT ATTTTCTT
4201 GACTTCTTTT ATATAGTAAT AAAAGTTATT TTGAAAAAA AAAAAAAA
4251 AAAA

```

BLAST Results

Entry HSG20626 from database EMBL:
human STS A005227.
Score = 860, P = 3.0e-32, identities = 176/181

Medline entries

89030633:
The structure of cytochrome b561, a secretory vesicle-specific electron transport protein.

Peptide information for frame 2

ORF from 74 bp to 931 bp; peptide length: 286
Category: strong similarity to known protein
Classification: unset

```

1 MAMEGYRRFL ALLGSALLVG FLSVIFALVW VLHYREGLGW DGSALFENWH
51 PVLMTVTFVF IQGIAIIVYR LPWTWKCSKL LMKSIHAGLN AVAILAIIS
101 VVAVFENHNH NNIANMYS LH SWGLIAVIC YLLQLSGFS VFLLPWAPLS
151 LRAFLMPIHV YSGIVIFGTV IATALMGLTE KLIFSLRDPY YSTFPPEGVF
201 VNTLGLLILV FGALIFIVT RPQWKRPKEP NSTILHPNGG TEQGARGSMP
251 AYSGNMMDKS DSELNNEVAA RKRNLALDEA QORSTM

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_7e22, frame 2

SWISSPROT:C561_SHEEP CYTOCHROME B561 (CYTOCHROME B-561)., N = 1, Score

342

DKFZphfbr2_7j4

group: brain derived

DKFZphfbr2_7j4 encodes a novel 233 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit

Sequenced by GBF

Locus: unknown

Insert length: 1050 bp

Poly A stretch at pos. 1027, polyadenylation signal at pos. 1007

```
1 GGGGACACAA AGGGGTGGTC ACCCTGCCCT CACCTTGACC TGTAAGTTGC
51 CTAGGACAGT GGCCTGGTCC CAGGGGCTGT TGTGGGGAGT TGAAGAACAC
101 CCTGGCCTCC TCCATCATGT CGGCCAAGAG GGCAGAATTG AAGAAAACAC
151 ATCTGTGCAA GAACTACAAG GCAGTTTGCC TGGAATTGAA GCCAGAGCCG
201 ACCAAAAACAT TTGATTACAA AGCAGTTAAA CAAGAAGGGC GGTTTACCAA
251 AGCAGGAGTG ACACAGGACC TAAAGAAATGA ACTCAGGGAA GTGAGAGAAG
301 AGCTCAAGGA GAAAATGGAG GAGATAAAAC AGATAAAGGA TCTAATGGAC
351 AAGGATTTTG ATAACTTCA CGAATTTGTG GAAATTATGA AGGAAATGCA
401 GAAAGATATG GATGAGAAGA TGGACATTTT AATAAATACA CAGAAGAAGT
451 ATAAGCTTCC CCTTAGAAGA GCACCAAAGG AGCAGCAGGA ACTCAGGCTG
501 ATGGGAAAGA CTCACAGAGA ACCACAGCTC AGGCCCAAGA AAATGGATGG
551 AGCCAGTGGA GTCAATGGAG CACCCTGTGC TCTTCACAAG AAGACGATGG
601 CACCACAAAA AACAAAACAG GGCTCACTGG ATCCCCTTCA TCACTGTGGG
651 ACCTGCTGCG AGAAATGTTT GTTGTGTGCT CTAAAGAACA ACTACAATCG
701 GGGGAACATT CCTTCAGAGG CCTCAGGCCT TTACAAAGGT GGAGAGGAGC
751 CAGTGACCAC CCAACCTTCT GTGGGCCACG CTGTGCCTGC CCCAAAGTCC
801 CAGACTGAGG GAAGGTGAAG CTTAACTGCC AGCTTGAAAT GAGAGTAAAG
851 AAGATACAGA GCAAACAGTG TTTCAGAAAC TGTCTGCCCT TGGGTGTGAT
901 TCTTTGGCTT CAATTTGAAG GAGGAGGAAT GATGGGATTT CATATTTTAT
951 TTCACACCAG TTCCTCCTTG TTTCATCTCT TTGCTAAGCT GGCTGCTTCT
1001 ACCATCTAAT AAATAATTGG CCAAGTTAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3
-----ORF from 117 bp to 815 bp; peptide length: 233
Category: putative protein

```
1 MSAKRAELKK THLCKNYKAV CLELKPEPTK TFDYKAVKQE GRFTKAGVTO
51 DLKNELREVR EELKEKMEEI KQIKDLMDKD FDKLHEFVEI MKEMQKDMDE
101 KMDILINTQK NYKLPLRRAP KEQQLRLMG KTHREPQLRP KRMDGASGVN
151 GAFCALHKKT MAPQKTKQGS LDPLHHCCTC CEKCLLCALK NNYNRGNIPS
201 EASGLYKGGE EPVTTQPSVG HAVPAPKSQT EGR
```

BLASTP hits

Entry JC2223 from database PIR:

major surface glycoprotein 3 - Pneumocystis carinii (fragment)

Score = 109, P = 3.5e-04, identities = 41/136, positives = 67/136

TREMBLNEW:PCP115C_1 product: "P115C"; Pneumocystis carinii mRNA for P115C, partial sequence., N = 1, Score = 109, P = 0.00024

HSPs:

```

Query:      14  CKN-YKAVCLCLKPEPTKTFDYKAVKQEGRFTKA-GVTQDLKNELREVREELKEKMEEIK 71
            CK  K  K  CEEL  + K VK+  TK  G  ++LK+++++  E  KE++E  K
Sbjct:      22  CKTELK KCYCELKKEADGLKVNDR-KVEICDDTKRDGCKELKDVKKELETFFEEL--K 78

Query:      72  QIKDLMDKDFDKLHEFVEIMKEMQKMDDEKMILINTQKNYKPLRRAPKEQQELRLMGK 131
            +KD+ D++ +K  E  +++E  D  D  K  +  +  +  YKL  +R  E  LR  +GK
Sbjct:      79  ALKDIDKDNCEKYEEKCILLEETNHD-DVKKNCVKLREGCYKLRKRVA-EDLLLRLALGK 136

Query:     132  THREPQLRPKKMDGAS 147
            +  +  K  D  S
Sbjct:     137  DVKNGECEKKMKDVCS 152

```

Report for DKFZphfbr2 7j4.3

```
[LENGTH]          233
[MW]               26533.95
[pI]               9.18
[PROSITE]          MYRISTYL          3
[PROSITE]          CK2_PHOSPHO_SITE      3
[PROSITE]          PKC_PHOSPHO_SITE      3
[KW]               All_Alpha
[KW]               LOW_COMPLEXITY      14.59 %
[KW]               COILED_COIL      13.73 %
```

```

SEQ      MSAKRAELKKTTHLCKNYKAVCLELKEPTKTFTDYKAVKQEGRFTKAGVTQDLKNELREV
SEG      .
PRD      cccchhhhhhhhhhhccchhhhhhhccccccccccccceeeccccccccccccchhhhhhhhhh
COILS    . . . . . CCCCCCCCCC

SEQ      EELKEKMEEIKQIKDLMKDQDFDLKHEFVEIMKEMQKMDKMDILINTQKNYKPLRRAP
SEG      xxxxxxxxxx . . . . . xxxxxxxxxxxxxxxx . . . . .
PRD      hhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhchhhhhhhhhcccccccccc
COILS    CCCCCCCCCCCCCCCCCC

SEQ      KEQQELRLMGKTHREPOLRPKKMDGASGVNGAPCALHKKTMAPQKTKQGSLDPLHHGCTC
SEG      .
PRD      hhhhhhhhhccccccccccccccccccccccccchhhhhhhcccccccccccccccccc
COILS    .

SEQ      CEKCLLCALKNNYNRGNIPSEASGLYKGGEPTVTQPSVGHAVPAKKSQTEGR
SEG      .
PRD      chhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .

```

Prosite for DKF2phfbr2 7i4.3

PS00005	2->5	PKC_PHOSPHO_SITE	PDOC00005
PS00005	108->111	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	179->183	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00008	151->157	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2 7j4.3)

DKFZphfbr2_82c20

group: transmembrane protein

DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with very weak similarity to C. elegans cosmid D1007.

The novel protein contains 7 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans D1007.5 ;
membrane regions: 7
Summary DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with similarity to a hypothetical C.elegans protein.

similarity to C.elegans D1007.5

complete cDNA (Bp 1-100 GC ritch), complete cds,
potential start at Bp 128 matches Kozak consensus PyNnatgG,
EST hits, localisation? primer B of STS doesn't match perfect!
TRANSMEMBRANE 7

Sequenced by DKFZ

Locus: /map="109.9 cR from top of Chr1 linkage group"???

Insert length: 1804 bp

Poly A stretch at pos. 1794, no polyadenylation signal found

```
1 CGGCGGGAGC GCGCGGCTGA TACCCGGGAC TGGGCTGCGG CGGTAGTCC
51 TCTCCCGGCC GCCGTCGCCT CCGACATATT GCTCGCAGGA GCTGCGGCGG
101 CGAAGCGGAG AGCACCAGGG GGAGGAGATG GGAGGACGAA GAGGTCCCAA
151 CAGGACATCT TACTGTGCAA ATCCGCTCTG TGAGCCGGGA TCCTCGGGGG
201 GCTCTAGTGG AAGCCACACT TCCAGTGCAT CGGTGACCAG TGTTCGTTCC
251 CGCACCAGGA GAGTTCTTGG AACAGGCCTC TCCAGCCCTC CTCTGGCCAC
301 CCAAACTGTT GTGCCTCTAC AGCACTGCAA GATCCCCGAG CTGCCAGTCC
351 AGGCCAGCAT TCTGTTTGAG TTGCAGCTCT TCTTCTGCCA GCTCATAGCA
401 CTCTTCGTCC ACTACATCAA CATCTACAAG ACAGTGTGGT GGTATCCACC
451 TTCCCACCCA CCTCCACACA CCTCCCTGAA CTTCCATCTG ATCGACTTCA
501 ACTTGCTGAT GGTGACCACC ATCGTTCTGG GCCGCCGCTT CATTTGGTCC
551 ATCGTGAAGG AGGCCTCTCA GAGGGGGAAG GTCTCCCTCT TTCGCTCCAT
601 CCGTGTGTTT CTCACCTGCT TCACCGTTCT CACGGCAACA GGCTGGAGTC
651 TGTGCCGATC CCTCATCCAC CTCTTCAGGA CTTACTCCTT CCTGAACCTC
701 CTGTTCTCTT GCTATCCGTT TGGGATGTAC ATTCCGTTCC TGCAGCTGAA
751 TTGCGACCTC CGCAAGACAA GCCTCTTCAA CCACATGGCC TCCATGGGGC
801 CCCGGGAGGC GGTCACTGGC CTGGCAAAGA GCCGGGACTA CCTCCTGACA
851 CTGCGGGAGA CGTGGGAAGCA GCACACAAGA CAGCTGTATG GCCCGGACGC
901 CATGCCCAAC CATGCCCTGCT GCCTGTCAAC CAGCCTCATC CGCAGTGAGG
951 TGGAGTTCCT CAAGATGGAC TTCAACTGGC GCATGAAGGA AGTGCTCGTC
1001 AGCTCCATGC TGAGCGCCTA CTATGTGGCC TTTGTGCCTG TCTGGTTCGT
1051 GAAGAACACA CATTACTATG ACAAGCGCTG GTCCTGTGAA CTCTTCTGTC
1101 TGGTGTCATC CAGCACCTCC GTGATCCTCA TGCAGCACTT GCTGCCCTGCC
1151 AGCTACTGTG ACCTGCTGCA CAAGGCCGCC GCCCATCTGG GCTGTTGGCA
1201 GAAGGTGGAC CCAGCGCTGT GCTCCAACGT GCTGCAGCAC CCGTGGACTG
1251 AAGAAATGCAT GTGGCCGAG GCGGTGCTGG TGAAGCACAG CAAGAACGTC
1301 TACAAAGCCG TAGGCCACTA CAACGTGGCT ATCCCCTCTG ACGTCTCCCA
1351 CTTCCGCTTC CATTCTTTT TCAGCAAACC TCTGCGGATC CTCAACATCC
1401 TCCTGCTGCT GGAGGGCGCT GTCATTGTCT ATCAGCTGTA CTCCCTAATG
1451 TCCTCTGAAA AGTGGCACCA GACCATCTCG CTGGCCCTCA TCCTCTTCAG
1501 CAACTACTAT GCCTTCTTCA AGCTGCTCCG GGACCGCTTG GTATTGGGCA
1551 AGGCCTACTC ATACTCTGCT AGCCCCAGA GAGACCTGGA CCACCGTTTC
1601 TCCTGAGCCC TGGGGTCACC TCAGGGACAG CGTCCAGGCT TCAGCCAAGG
1651 GCTCCTTGCC AAGGGGCTGT TGGGTAGAAG TGGTGGTGGG GGGGACAAAA
1701 GACAAAAAAA TCCACCAGAG CTTTGTATTT TTGTTACGTA CTGTTTCTTT
1751 GATAATTGAT GTGATAAGGA AAAAAGTCCT ATTTTATATC TCCCAAAAAA
1801 AAAA
```

BLAST Results

Entry HS285343 from database EMBL:
human STS WI-17488.

Score = 1225, P = 1.3e-50, identities = 263/281

Medline entries

No Medline entry

Peptide information for frame 2

```

1  MGGRRGPNRT SYCRNPCEP GSSGGSSGSH TSSASVTSVR SRTSSSSGTG
51 LSSPPLATQT VVPLQCHKIP ELPVQASILF ELQLFFCQLI ALFVHYINIY
101 KTVWWYPPSH PPSHTSLNFH LIDFNLMVT TIVLGRRFIG SIVKEASQRG
151 KVSIFRSILL FLTRFTVLTG TGWSLCRSLI HLFRTYSFLN LLFLCYPFGM
201 YIPFLQLNCD LRKTSLEFNM ASMGPREAVS GLAKSRDYLL TLRETWKQHT
251 RQLYGPDAMP THACCLSPSL IRSEVEFLKM DFNWRMKEVL VSSMLSAYYV
301 AFVPVWFVKN THYYDKRWSC ELFLVLSIST SVILMQHLLP ASYCDLLHKA
351 AAHLGCWQKV DPALCSNVLQ HPWTEECMWP QGVLVKHKN VYKAVGHYNV
401 AIPSDVSHFR FHFFFSKPLR ILNILLLEG AVIVYQLYSL MSSEKWHQTI
451 SLALILFSNY YAFFKLLRDR LVLGKAYSYS ASPQRDLDR FS

```

ORF from 128 bp to 1603 bp; peptide length: 492

Category: similarity to unknown protein

Prosite motifs: LEUCINE ZIPPER (210-232)

LEUCINE_ZIPPER (210-232)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_82c20, frame 2

TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid D1007., N = 2, Score = 247, P = 4.6e-29

>TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid D1007.
Length = 512

HSPs:

Score = 247 (37.1 bits), Expect = 4.6e-29, Sum P(2) = 4.6e-29
Identities = 58/204 (28%), Positives = 102/204 (50%)

```

Query: 291 VSSMLSAYYVAFVPVWFVKNTHYYDKRWSCFLVLSISTSVILMQHLLPASVCDLLHKA 350
      +S ML +V F + ++ W C+L ++V ++ + + +L P +Y DLLH+A
Sbjct: 299 LSIMLPCIFVPFKTSQGIPOKILINEVWECQLAIVVGLTAFSLYVAYLSPLNYLDLLHRA 358

Query: 351 AAHLGCWQKVD-PAL----CSNVLQHPWTEECMWPQGVLVKHKN-VYKAVGHYNV---- 400
      A HLG W +++ P + + PW+E C++ G V+ Y+A ++
Sbjct: 359 AIHLGSHWHQIEGPRIGHTGSMSSAPTWSSEFLYNDGETVQMPDGRYRAKSSNSIRTVA 418

Query: 401 AIPSDVSHFRFHFFFSKPLRILNILLLEGAVIVYQLYSLMSSEKWHQTI SLALILFSNY 460
      A P H F KP ++NI+ E +I Q + L+ + W ++ L++F+NY
Sbjct: 419 AHPSSRHNTFFKVLKRPNNLINMCSFEFLIFIQFWMLVLTNDWQHIVTFVLLMFANY 478

Query: 461 YAFFKLLRDRVLGKAYSYSASPQRDL 487
      F KL +D+++L + Y S Q DL
Sbjct: 479 LLFAKLKDKIILSRIEPS---QEDL 502

```

Score = 178 (26.7 bits), Expect = 4.3e-21, Sum P(2) = 4.3e-21
Identities = 50/179 (27%), Positives = 90/179 (50%)

```

Query: 262 HACCLSPSLIRSEVEFLKMDFNWRMKEVLVSSMLSAYYVAFVPVWFV--KNTHYYDKR-- 317
      H C SP+ IR E++ L D R+K+ + + +A+ +P FV K + ++
Sbjct: 262 HMCSDSPAQIREIQVLIDDLVLRVKKSI FAGVSTAFSLSIMLPCIFVPFKTSQGIPOKIL 321

Query: 318 ---WSCEFLVLSISTSVILMQHLLPASVCDLLHKA AAHLGCWQKVD-PAL----CSNV 368
      W C+L ++V ++ + + +L P +Y DLLH+AA HLG W +++ P + +
Sbjct: 322 INEVWECQLAIVVGLTAFSLYVAYLSPLNYLDLLHRAAHLGSHWHQIEGPRIGHTGSMSS 381

Query: 369 LQHPWTEECMWPQGVLVKHKN-VYKAVGHYNV-AIPSDVSHFRFHFFFSKPLRILNILL 426
      PW+E C++ G V+ Y+A ++ + + R + FF K LR N L+
Sbjct: 382 APTPWSEFLYNDGETVQMPDGRYRAKSSNSIRTVA AHPSSRHNTFF-KVLKRPNNLI 440

```

Report for DKF2phfbr2 82c20.2

[PROSITE]	LEUCINE_ZIPPER	1	
[PROSITE]	AMIDATION	2	
[PROSITE]	MYRISTYL	5	
{PROSITE}	CAMP_PHOSPHO_SITE	2	
[PROSITE]	CK2_PHOSPHO_SITE	3	
[PROSITE]	GLYCOSAMINOGLYCAN	1	
[PROSITE]	PKC_PHOSPHO_SITE	5	
[PROSITE]	ASN_GLYCOSYLATION	1	
[KW]	TRANSMEMBRANE	7	
[KW]	LOW_COMPLEXITY	8.74	%

```
SEQ      ILNILLLEGAIVYQLYSLSMSEKWHQTISLALILFSNYAYFFKLLRDLVLGKAYSYS
SEG      xxxxxxxx.....
PRD      hhhhhhhhheeeeeeccccccccccccccccccccccccccccccccccccccccc
MEM      mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm
```

```

SEQ   ASPQRDLDRFS
SEG   .....
PRD   ccchhhhhccc
MEM   .....

```

Prosites for DKFZphfbr2_82c20.2

PS00001	8->12	ASN_GLYCOSYLATION	PDOC00001
PS00002	47->51	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	212->216	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	316->320	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	443->446	PKC_PHOSPHO_SITE	PDOC00005
PS00006	241->245	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	28->34	MYRISTYL	PDOC00008
PS00008	48->54	MYRISTYL	PDOC00008
PS00008	231->237	MYRISTYL	PDOC00008
PS00009	2->6	AMIDATION	PDOC00009
PS00009	134->138	AMIDATION	PDOC00009
PS00029	168->190	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_82c20.2)

DKFZphfbr2_82e17

group: transmembrane protein

DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with very weak similarity to C. elegans cosmid R01B10.

The novel protein contains 6 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans "R01B10.5" ;
membrane regions: 6
Summary DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with
similarity to a hypothetical C.elegans protein.

similarity to C.elegans "R01B10.5"

complete cDNA, EST HS763158 extends the sequence, complete cds, EST
hits
six potential transmembrane domains

Sequenced by DKFZ

Locus: /map="779_C_?; 818_A_1; 877_C_1; 734_C_12; 760_E_11; 171.7 cR from top of Chr14 linkage
group"

Insert length: 1618 bp

Poly A stretch at pos. 1608, polyadenylation signal at pos. 1588

```

1 CTGATCTAGT GCTTCTCGAA AAAAACCTTC AGGCGGCCCA TGGCTGTCGA
51 TATTCAACCA GCATGCCTTG GACTTTATTG TGGGAAGACC CTATTATTTA
101 AAAATGGCTC AACTGAAATA TATGGAGAAT GTGGGGTATG CCCAAGAGGA
151 CAGAGAACGA ATGCACAGAA ATATTGTCAG CCTTGCACAG AATCTCCTGA
201 ACTTTATGAT TGGCTCTATC TTGGATTTAT GGCAATGCTT CCTCTGGTTT
251 TACATTGGTT CTTTATTGAA TGGTACTCGG GGAAAAAGAG TTCCAGCGCA
301 CTTTTCCAAC ACATCACTGC ATTATTGAA TGCAGCATGG CAGCTATTAT
351 CACCTTACTT GTGAGTGATC CAGTTGGTGT TCTTTATATT CGTTCATGTC
401 GAGTATTGAT GCTTTCTGAC TGGTACACGA TGCTTTACAA CCCAAGTCCA
451 GATTACGTTA CCACAGTACA CTGTACTCAT GAAGCCGTCT ACCCACTATA
501 TACCATTGTA TTTATCTATT ACGCATTTCT CTGTGTATTA ATGATGCTGC
551 TCCGACCCTC TCTGGTGAAG AAGATTGCAT GTGGGTTAGG GAAATCTGAT
601 CGATTTAATA GTATTATGTC TGCACTTTAC TTCTTCCCAA TTTTAACCGT
651 GCTTCAGGCA GTTGGTGGAG GCCTTTTATA TTACGCCTTC CCATACATTA
701 TATTAGTGTT ATCTTTGGTT ACTCTGGCTG TGTACATGTC TGCTTCTGAA
751 ATAGAGAAGT GCTATGATCT TCTGGTCAGA AAGAAAAGAC TTATTGTTCT
801 CTTCAGCCAC TGGTTACTTC ATGCCCTATG AATAATCTCC ATTTCCAGAG
851 TGGATAAAGT TGAGCAAGAT TTGCCCTTTT TGGCTTTGGT ACCTACACCA
901 GCCCTTTTAT ACTTGTTTAC TGCAAAATTT ACCGAACCTT CAAGGATACT
951 CTCAGAAGGA GCCAATGGAC ACTGAGTGTA GACATGTGAA ATGCCAAAAA
1001 CCTGAGAAGT GCTCCTAATA AAAAAGTAAA TCAATCTTAA CAGTGTATGA
1051 GAACTATTCT ATCATATATG GGAACAAGAT TGTCAGTATA TCTTAATGTT
1101 TGGGTTTGTC TTTGTTTTGT TTATGGTTAG ACTTACAGAC TTGGAAATG
1151 CAAAACCTCT TAATACTCTG TTACACAGGG TAATATTATC TGCTACACTG
1201 GAAGGCCGCT AGGAAGCCCT TGCTTCTCTC AACAGTTCAG CTGTTCTTTA
1251 GGGCAAAATC ATGTTTCTGT GTACCTAGCA ATGTGTTCCC ATTTTATTAA
1301 GAAAAGCTTT AACACGTGTA ATCTGCAGTC CTTAACAGTG GCGTAATTGT
1351 ACGTACCTGT TGTGTTTCAG TTTGTTTTTC ACCTATAATG AATTGTAAAA
1401 ACAAAACATC TTGTGGGGTC TGATAGCAAA CATAGAAATG ATGTATATTG
1451 TTTTTTGTTA TCTATTTATT TTCATCAATA CAGTATTTTG ATGTATTGCA
1501 AAAATAGATA ATAATTTATA TAACAGGTTT TCTGTTTATA GATTGGTTCA
1551 AGATTGTTT GGATTATTGT TCCTGTAAAG AAAACAATAA TAAAAAGCTT
1601 ACCTACATAA AAAAAAAA

```

BLAST Results

Entry HS981146 from database EMBL:
human STS WI-6253.
Length = 208
Minus Strand HSPs:
Score = 1040 (156.0 bits), Expect = 1.9e-40, P = 1.9e-40

Identities = 208/208 (100%), Positives = 208/208 (100%), Strand = Minus
/ Plus

Entry HSG20716 from database EMBL:

human STS A006D06.

Length = 195

Minus Strand HSPs:

Score = 975 (146.3 bits), Expect = 1.8e-37, P = 1.8e-37

Identities = 195/195 (100%), Positives = 195/195 (100%), Strand = Minus
/ Plus

Medline entries

No Medline entry

Peptide information for frame 1

```

1 MAVDIQPAQL GLYCGKTLF KNGSTEIYGE CGVCPRGQRT NAQKYCQPCT
51 ESPELYDWLY LGFMAMLPV LHWFFIEWYS GKKSSSALFQ HITALFECM
101 AAIITLLVSD PVGVLYIRSC RVLMLSDWYT MLYNPSPDYV TTVHCTHEAV
151 YPLYTIVFIY YAFCLVLMML LRPLLVKKIA CGLGKSDRFK SIYAALYFFP
201 ILTVLQAVGG GLLYYAFPIY ILVLSLVTIA VYMSASEIEN CYDLLVRKKR
251 LIVLFSHWLL HAYGIISISR VDKLEQDLPL LALVPTPALF YLFTAKFTEP
301 SRILSEGANG H

```

ORF from 40 bp to 972 bp; peptide length: 311

Category: similarity to unknown protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82e17, frame 1

TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid
R01B10., N = 1, Score = 399, P = 1.4e-36

>TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid R01B10.
Length = 670

HSPs:

Score = 399 (59.9 bits), Expect = 1.4e-36, P = 1.4e-36
Identities = 95/280 (33%), Positives = 152/280 (54%)

```

Query:      2 AVDIQPAQLGLYCGKTLFKN-----GSTEIYGE CGVCPRGQRTNAQKYCQP 49
            A IQP+CLG +CG+T+L N GST + CG C G R NA C+ C
Sbjct:     292 ASTIQPSCLG-FCGRTVLVGNYSDEVEATTTAAGSTSL-SRCGPCSFGYRNAMSICES 349

Query:      50 TESPELYDWLYLGFMAMLPVLVHWWFFIEWYSGKKSSSALFQ---HITALFECMAAIITL 106
            + YDW+YL F+A+LPL+LH FI + K + ++ ++ + E +A +I +
Sbjct:     350 DTPLQPYDWMYLLFIALLPVLLHMQFIR-IARKYCRTRYEVSEYLCVILENVIACVIAV 408

Query:      107 LVSDPVGVLVYIRSCRVLMMLSDWYTMLYNPSPDYVTTVHCTHEAVYPLYTIVFIYYAFCLV 166
            L+ P ++ C + +WY YNP Y T+ CT+E V+PLY+I FI++ +
Sbjct:     409 LIYPPREFTFLLNGCSKTDIKEWYPACYNPRIGYTKTMRCTYEVVFPLYSITFIHHLILIG 468

Query:      167 LMMLLRPLLVKKIA CGLGKSDRFKSIYAALYFFPILTVLQAVGGGLLYYAFPIILVLSL 226
            +++LR L + L K+ K YAA+ PIL V+ AV G+++Y FPYI+L+ SL
Sbjct:     469 SILVLRSTLYCVL---LYKTYNGKPFYAAIVSVPI LAVIHAVLSGVVFYTFPYILLIGSL 525

Query:      227 VTLAVYMSASEIENCYDLLVR----KKRLIVLFSHWLLHAYGIISI 268
            + +++ ++VR LI L L+ ++G+I+I
Sbjct:     526 WAMCFHLALEGKRPLKEMIVRIATSPTHLIFLSITMLMLSFGVIAI 571

```

Pedant information for DKFZphfbr2_82e17, frame 1

Report for DKFZphfbr2_82e17.1

[PROSITE]	AMIDATION	1	
[PROSITE]	MYRISTYL	3	
[PROSITE]	CAMP_PHOSPHO_SITE	1	
[PROSITE]	CK2_PHOSPHO_SITE	3	
[PROSITE]	PKC_PHOSPHO_SITE	4	
[PROSITE]	ASN_GLYCOSYLATION	1	
[KW]	TRANSMEMBRANE	6	
[KW]	LOW_COMPLEXITY	7.72	%

Prosites for DKFZphfbr2_82e17.1

(No Pfam data available for DKFZphfbr2_82e17.1)

DKFZphfbr2_82e4

group: signal transduction

DKFZphfbr2_82e4 encodes a novel 473 amino acid protein with strong similarity to the calmodulin-binding proteins.

The novel protein is similar to human and rat Ca²⁺/calmodulin-dependent protein kinase (EC 2.7.1.123), rat calmodulin-binding protein, calmodulin binding protein kinase of *Fugu rupies* and *Rattus norvegicus* calcium/calmodulin-dependent protein kinase I. Calmodulin is the archetype of the family of calcium-modulated proteins of which nearly 20 members have been found. Calmodulin is involved in regulation of growth and cell cycle as well as in signal transduction and the synthesis and release of neurotransmitters. The novel protein seems to be involved in calmodulin-mediated pathways in human neuronal cells.

The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

strong similarity to calmodulin-binding proteins

complete cDNA, complete cds, EST hits
splice variant in comparison to rat I56542
ESTs HS2254543/HS1141907 define splice variant
see also DKFZphfbr2_82g20 unspliced form

Sequenced by DKFZ

Locus: /map="200.5 cR from top of Chr3 linkage group"

Insert length: 2923 bp

Poly A stretch at pos. 2913, polyadenylation signal at pos. 2890

```
1 ATGCTGGAGG TTCGCTAGCC GAAGCGGCTG CATCTGGCGC CGCGTCTGCC
51 CCGCGTGTCT GGAGCGGATT CTGCCC GCCG CCTCGGCGCC
101 CCGCTGAGCC CGCGATCACT TCCTCCCTGT GACCAACCGG CGCTGCAGGT
151 TAGAGCCTGG CAATGCCGTT TGGGTGTGTG ACTCTGGGTG ACAAGAGAA
201 CTATAACCCAG CCATCGGAGG TGAATGACAG ATATGATTGT GGACAGGTCA
251 TCAAGACTGA GGAGTTTGTG GAAATCTTCC GGGCCAAAGG CAAGACGACA
301 GGCAAGCTGC ACACCTGCAA GAAGTTCCAG AAGCGGGACG GCCGCAAGGT
351 GCGGAAAGCT GCCAAGAACG AGATAGGCAT CCTCAAGATG GTGAAGCATC
401 CCAACATCCT ACAGCTGGTG GATGTGTTTG TGACCCGCAA GGAGTACTTT
451 ATCTTCCCTGG AGCTGGCCAC GGGGAGGGAG GTGTTTGACT GGATCCTGGA
501 CCAGGGCTAC TACTCGGAGC GAGACACAAG CAACGTGGTA CGGCAAGTCC
551 TGGAGGCCGT GGCCTATTGT CACTCACTCA AGATCGTGCA CAGGAATCTC
601 AAGCTGGAGA ACCTGGTTTA CTACAACCGG CTGAAGAAGT CGAAGATTGT
651 CATCAGTGAC TTCCATCTGG CTAAGCTAGA AAATGGCCTC ATCAAGGAGC
701 CCTGTGGGAC CCCCAGTAT CTGGGCAACC CACCTTTCTA TGAGGAGGTG
751 GAAGAAGATG ATTATGAGAA CCATGATAAG AATCTCTTCC GCAAGATCCT
801 GGCTGGTGAC TATGAGTTTG ACTCTCCATA TTGGGATGAT ATTCGACAGG
851 CAGCCAAAGA CCTGGTCACA AGGCTGATGG AGGTGGAGCA AGACCAGCGG
901 ATCACTGCGC AAGAGGCCAT CTCCCATGAG TGGATTCTTG GCAATGCTGC
951 TTCTGATAAG AACATCAAGG ATGGTGTCTG TGCCAGATT GAAAAGAAGT
1001 TTGCCAGGGC CAAGTGAAG AAGGCTGTCC GAGTGACCAC CCTCATGAAA
1051 GCGCTCCGGG CACCAGAGCA GTCCAGCACG GCTGCAGCCC AGTCGGCCTC
1101 AGCCACAGAC ACTGCCACCC CCGGGGCTGC AGGTGGGGCC ACAGCTGCAG
1151 CTGCGAGTGG AGCTACCTCA GCCCCTGAGG GTGATGCTGC TCGTGTGCA
1201 AAGAGTGATA ATGTGGCCCC CGCAGACCGT AGTGCCACCC CAGCCACAGA
1251 TGGAAGTGCC ACCCCAGCCA CTGATGGCAG TGTCACCCCA GCCACCGATG
1301 GAAGCATCAC TCCAGCCACT GATGGGAGTG TCACCCACGC CACTGACAGG
1351 AGCGCTACTC CAGCCACTGA TGGGAGAGCC ACACCAGCCA CAGAAGAGAG
1401 CACTGTGCCC ACCACCCAAA GCAGTGCCAT GCTGGCCACC AAGGCAGCTG
1451 CCACCCCTGA GCCGGCTATG GCCCAGCCGG ACAGCACAGC CCCAGAGGGC
1501 GCCACAGGCC AGGCTCCACC CTCTAGTAAA GGGGAAGAGG CTGCTGGTTA
1551 TGCCCAAGGAG TCTCAAAGGG AGGAGGCCAG CTGAGTAGGC AGCCTGGTGA
1601 GGGGGGGCAG GGGATGGGCA GGAGGGTGGG AGAGTGGATG AGGGGCTTCT
1651 CACTGTACAT AGAGTCACTG GCATGATGCC CTCGCTCCCC CATGCCCCCA
1701 CATCCCACTG GGGCATAACT AGGGGTGACG GGAGAGCAGT CTCGTCTCCT
1751 GTGTGTATGT GTGTGAGTGG TGGGCAGGCC AGTGGCAGGG CCGGCCCCAG
1801 CCCCTGCATG GATTCCCTGT GGCTTTTCTG TCTTTTGCTA GCTTCACCAG
1851 TTTCTGTTCC TTGTGGGATG CTGCTCTAGG GATACTCAGG GGGCTCCTGC
1901 TCTCCTTCCC CTTCCTTCTT TGCTCACCA TTCCCTTAGG CAGGCCCTGC
1951 AGGTCCACCA CTCTCCAGG CCCTAAACTT GGGCGGCCTT GCCCTGAGAG
2001 CTGGTCTCTC AGCGAGGCCG GTGCAGCGGT CTTAGGCTCC TGACATGAA
2051 GGTGTGTGTC TGTGGTGTGT GGGCTGCTCT AGGAGCAGAT ACAGGCTGGT
2101 ATAGAGGATG CAGAAAGGTA GGCAGTATG TTTAAGTCCA GACTTGGCAC
2151 ATGGCTAGGG ATACTGCTCA CTAGCTGTGG AGGTCTCTAG GAGTGGAGAG
2201 AATGAGTAGG AGGGCAGAAG CTTCCATTTT TGTCTTCTT AAGACCTCTG
```

```

2251 TATTTGTGTT ATTTCTGCCC TTTCAGAGTC CTGCAGTGGG CTGCCCTGTA
2301 CCCTGAACCT CATGAGCCTC TAAGGGAAAAG GAGGAACAAT TAGGACGTGG
2351 CAATGAGACC TGGCAGGGCA GAGTACAAGC CCAGCACCCA GTGTCCCAGC
2401 CTTACTGGGT CCTTACCCTG GGCCAAACAG GGAGGGCTGA TACCTCCTTG
2451 CTCTTCCCTAG ATGCCCACCT CCTACAATCT CAGCCACACAA GTCCTCTCCA
2501 CCCTAGGGGG CTGCTGCAT GGCAATAACT CATAATCTGA TTGGAGGTT
2551 TGCCCTTTAC AGGGGAGAT TTTCTGCTCA GTTCAACAAT GAAATGAAGA
2601 GGAACCTCCCT CTTTCTACAG CTCACTTCTA TCAGAGGGCC AGGTGCCTCA
2651 GAGCCACATT GAGTTGCTTT TTCTGGGATG AGGAAGTAGG GTTAAACTCC
2701 CCAGTTTCTT GAGGGAGGCT CCTGACAGGT GCCCTTTGTC AGACCCTACC
2751 ACAGCTCTGA TAGGCAGCCA CATTGGTCCT CGCCCTTGCT CGGCACCTCCG
2801 TGGTGGTCTT GCCCTTCTCC CTGCATGCCT GTGGGTCTGC TCTGGTGTGT
2851 GAAGGTCCGT GGGTTAACTG TGTGCCTACT GAACCTGGCA AATAAACATC
2901 ACCCTGCAAA GCCAAAAAAA AAA

```

BLAST Results

Entry HS452352 from database EMBL:
human STS WI-15318.
Length = 350
Minus Strand HSPs:
Score = 1547 (232.1 bits), Expect = 5.2e-63, P = 5.2e-63
Identities = 331/348 (95%), Positives = 331/348 (95%), Strand = Minus /
P1

Medline entries

94110847:
J Neurosci 1994 Jan;14(1):1-13
IG5: a calmodulin-binding, vesicle-associated, protein
kinase-like protein enriched in forebrain neurites.
Godbout M, Erlander MG, Hasel KW, Danielson PE, Wong KK, Battenberg EL,
Foye PE,
Bloom FE, Sutcliffe JG

Peptide information for frame 1

```

1 MPFGCVTLGD KKNYNQPSV TDRYDLGQVI KTEEFCEIFR AKDKTTGKLH
51 TCKKFQKRDG RKVRKAAKNE IGILKMVKHP NILQLVDVVF TRKEYFIFLE
101 LATGREVFDFW ILDQGYYSER DTSNVVRQVL EAVAYLHSLK IVHRNLKLEN
151 LVYYNRLKNS KIVISDFHLA KLENGLIKEP CGTPEYLGNP PFYEEVEEDD
201 YENHDKNLFR KILAGDYEDF SPYWDDISQA AKDLVTRLME VEQDQRITAE
251 EASHHEWISG NAASDKNIKD GVCAQIEKNF ARAKWKKA VR VTTLMKRLRA
301 PEQSSTAAAQ SASATDTATP GAAGGATAAA ASGATSAP EG DAARAASDN
351 VAPADRSATP ATDGSATPAT DGSVTPATDG SITPATDGSV TPATDRSATP
401 ATDGRATPAT EESTVPTTQS SAMLATKAAA TPEPAMAQPD STAPEGATGQ
451 APPSSKGEEA AGYAQESQRE EAS

```

ORF from 163 bp to 1581 bp; peptide length: 473
Category: strong similarity to known protein

BLASTP hits

Entry S50193 from database PIR:
Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - rat
Length = 374
Score = 371 (130.6 bits), Expect = 2.2e-66, Sum P(2) = 2.2e-66
Identities = 74/176 (42%), Positives = 115/176 (65%)

Entry S57347 from database PIR:
Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - human
Length = 370
Score = 369 (129.9 bits), Expect = 4.6e-66, Sum P(2) = 4.6e-66
Identities = 74/176 (42%), Positives = 114/176 (64%)

Alert BLASTP hits for DKFZphfbr2_82e4, frame 1

PIR:I56542 calmodulin-binding protein - rat, N = 2, Score = 1246, P =
4e-228

TREMBLNEW:FRU010348_3 product: "calmodulin binding protein kinase";
Fugu rubripes UBE1-like gene, PRGFR2 gene and gene encoding calmodulin
binding protein kinase, clone 168J21, N = 2, Score = 846, P = 2.6e-139

TREMBL:RNPRKI_1 product: "protein kinase I"; Rattus norvegicus
calcium/calmodulin-dependent protein kinase I mRNA, complete cds., N =
2, Score = 364, P = 5.1e-63

>PIR:I56542 calmodulin-binding protein - rat
Length = 504

HSPs:

Score = 1246 (186.9 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
Identities = 255/289 (88%), Positives = 259/289 (89%)

Query: 188 GNPPFFYEVEEDDYENHDKNLFKILAGDYEFDSPYWDISQAADLVTRLMEVEQDQRI 247
GNPPFFYEVEEDDYENHDKNLFKILAGDYEFDSPYWDISQAADLVTRLMEVEQDQRI
Sbjct: 216 GNPPFFYEVEEDDYENHDKNLFKILAGDYEFDSPYWDISQAADLVTRLMEVEQDQRI 275

Query: 248 TAEAAISHEWISGNAASDKNIKDGVCQAIEKNFARAKWKKAVRVTTLMKRLRAPEQSSTA 307
TAEAAISHEWISGNAASDKNIKDGVCQAIEKNFARAKWKKAVRVTTLMKRLRAPEQS TA
Sbjct: 276 TAEAAISHEWISGNAASDKNIKDGVCQAIEKNFARAKWKKAVRVTTLMKRLRAPEQSGTA 335

Query: 308 AAQSASATDTATPGAAGGATAAAASGATSAP-----GDAARAASDNVAPADRSAT 359
A +D ATPGAAGGA AAAA GA A GDA AAKSD++A ADRSAT
Sbjct: 336 AT-----SDAATPGAAGGAVAAAGGAAPASGASATVGTGGDAGCAAKSDDMASADRSAT 390

Query: 360 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQ 419
PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVP Q
Sbjct: 391 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPAAQ 450

Query: 420 SSAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS 473
SSA A KAAATPEPA+AQPDSTA EGATGQAPPSSKGEEA G AQESQR E S
Sbjct: 451 SSAAPAAKAAATPEPAVAQPDSTALEGATGQAPPSSKGEEATGCAQESQRVETS 504

Score = 978 (146.7 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
Identities = 186/187 (99%), Positives = 187/187 (100%)

Query: 1 MPFGCVTLGDKKNYNQPSVTDRLGQVQIKTEEFCEIFRAKDKTGKLTCKKFQKRDG 60
MPFGCVTLGDKKNYNQPSVTDRLGQV+KTEEFCEIFRAKDKTGKLTCKKFQKRDG
Sbjct: 1 MPFGCVTLGDKKNYNQPSVTDRLGQVQVQIKTEEFCEIFRAKDKTGKLTCKKFQKRDG 60

Query: 61 RKVRKAAKNEIGILKMKVHPNQLQVDFVTRKEYFIFLELATGREVFDWILDQGYYSER 120
RKVRKAAKNEIGILKMKVHPNQLQVDFVTRKEYFIFLELATGREVFDWILDQGYYSER
Sbjct: 61 RKVRKAAKNEIGILKMKVHPNQLQVDFVTRKEYFIFLELATGREVFDWILDQGYYSER 120

Query: 121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180
DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP
Sbjct: 121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180

Query: 181 CGTPEYL 187
CGTPEYL
Sbjct: 181 CGTPEYL 187

Pedant information for DKFZphfbr2_82e4, frame 1

Report for DKFZphfbr2_82e4.1

[LENGTH] 473
[MW] 51208.89
[pI] 5.30
[HOMOL] PIR:I56542 calmodulin-binding protein - rat 0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 2e-26
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 2e-26
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YDL101c] 8e-26
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YCL024w] 5e-24
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 7e-23
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
7e-23
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 1e-21
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 1e-21

[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-16
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 3e-16
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 1e-15
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 3e-14
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YCR073c] 6e-11
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNR031c] 8e-11
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YJL095w] 2e-09
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 5e-07
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 8e-07
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 1e-05
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR523c] 2e-04
[FUNCAT] c energy conversion [M. genitalium, MG109] 3e-04
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[BLOCKS] BL00939F
[SCOP] dlgol_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus] 3e-62
[SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)] 5e-59
[SCOP] dlkoa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis] 1e-75
[SCOP] dlkoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har] 1e-72
[SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase 4e-65
[SCOP] dlirk_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens)] 2e-56
[SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu)] 4e-71
[SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Homo)] 1e-50
[SCOP] dlydre_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo)] 3e-70
[SCOP] dlfmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Homo)] 5e-49
[SCOP] dlcdkb_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su)] 2e-72
[SCOP] d2hcka3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [huma] 5e-46
[SCOP] dlcsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe] 9e-42
[SCOP] dljsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)] 1e-56
[SCOP] dlckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)] 9e-52
[EC] 2.7.1.38 Phosphorylase kinase 3e-29
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 8e-66
[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 2e-17
[EC] 2.7.1.117 Myosin-light-chain kinase 2e-38
[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 2e-17
[EC] 2.7.1.37 Protein kinase 6e-28
[PIRKW] phosphotransferase 8e-66
[PIRKW] nucleus 2e-24
[PIRKW] transferase 8e-30
[PIRKW] calcium 2e-27
[PIRKW] duplication 4e-19
[PIRKW] tandem repeat 2e-31
[PIRKW] phorbol ester binding 1e-16
[PIRKW] zinc 1e-16
[PIRKW] cell cycle control 2e-20
[PIRKW] serine/threonine-specific protein kinase 8e-66
[PIRKW] phospholipid binding 1e-16
[PIRKW] autophosphorylation 8e-66
[PIRKW] brain 1e-14
[PIRKW] heterotetramer 2e-16
[PIRKW] polymer 3e-29
[PIRKW] mitosis 2e-20
[PIRKW] magnesium 7e-22
[PIRKW] ATP 8e-66
[PIRKW] alternative initiators 1e-29

[PIRKW] phosphoprotein 8e-66
 [PIRKW] apoptosis 2e-31
 [PIRKW] glycoprotein 4e-19
 [PIRKW] skeletal muscle 3e-28
 [PIRKW] protein kinase 2e-28
 [PIRKW] testis 3e-28
 [PIRKW] signal transduction 1e-21
 [PIRKW] cAMP binding 1e-16
 [PIRKW] purine nucleotide binding 5e-25
 [PIRKW] structural protein 4e-19
 [PIRKW] calcium binding 3e-45
 [PIRKW] alternative splicing 3e-45
 [PIRKW] P-loop 5e-25
 [PIRKW] lipoprotein 2e-16
 [PIRKW] cardiac muscle 4e-19
 [PIRKW] muscle 3e-28
 [PIRKW] myristylation 2e-16
 [PIRKW] EF hand 5e-29
 [PIRKW] cell division 2e-38
 [PIRKW] calmodulin binding 8e-66
 [PIRKW] smooth muscle 7e-31
 [SUPFAM] fibronectin type III repeat homology 7e-31
 [SUPFAM] immunoglobulin homology 7e-31
 [SUPFAM] ribosomal protein S6 kinase II 3e-26
 [SUPFAM] calcium-dependent protein kinase 5e-29
 [SUPFAM] AMP-activated protein kinase 7e-22
 [SUPFAM] protein kinase akt 1e-14
 [SUPFAM] protein kinase SPK1 3e-20
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-36
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 3e-45
 [SUPFAM] calmodulin repeat homology 5e-29
 [SUPFAM] protein kinase DUN1 2e-24
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-14
 [SUPFAM] death-associated protein kinase 2e-31
 [SUPFAM] myosin-light-chain kinase, nonmuscle 1e-29
 [SUPFAM] pleckstrin repeat homology 1e-14
 [SUPFAM] ankyrin repeat homology 2e-31
 [SUPFAM] protein kinase homology 8e-66
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 8e-36
 [SUPFAM] twitchin 1e-18
 [SUPFAM] protein kinase C zinc-binding repeat homology 1e-16
 [SUPFAM] titin 4e-19
 [SUPFAM] protein kinase cdrl 2e-20
 [SUPFAM] kinase-related transforming protein 2e-38
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 8e-66
 [SUPFAM] kinase interaction domain homology 2e-24
 [SUPFAM] protein kinase C mu 1e-16
 [PROSITE] AMIDATION 1
 [PROSITE] MYRISTYL 3
 [PROSITE] CK2_PHOSPHO_SITE 10
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 11
 [PFAM] Eukaryotic protein kinase domain
 [KW] All_Alpha
 [KW] 3D
 [KW] LOW_COMPLEXITY 7.40 %

SEQ MPFGCVTLGDKKNYNQPSVTDRLGQVIKTEEFCEIFRAKDKTTGKLHTCKKFQKRDG
 SEG
 1a06-CEETTTGGGCEEEEEECBCGGGGGEEEEETTTTCEEEEEEEEC---

SEQ RKVRKAAKNEIGILKMVKHPNQLQVDVFTVTRKEYFIFLELATGREVFDWILDQGGYYSER
 SEG
 1a06- -----HHHHHHHHHCCTTTBCCEEEEEETTEEEEECCCCCEEHHHHHHHTTTTBHH

SEQ DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYNNRLKNSKIVISDFHLAKLENGLIKEP
 SEG
 1a06- HHHHHHHHHHHHHHHHHHCCTTTTTTTEEECCCTTTTCEEECCCTTTTCHHHHHHCCC

SEQ CGTPEYLGNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSPLYWDDISQAADLVTRLME
 SEG
 1a06- HHHHHHHHCCTTTTTT-----THHHHHHHHCCCCCTTTTTTTTCHHHHHHHHHHCT

SEQ VEQDQRITAEAEISHEWISGNAASDKNIKDGVCQAIEKNFARAKWKKAVRVTTLMKRLRA
 SEG
 1a06- TTGGGCCCHHHHHHTTTTTTCCCCCHHHHHHHHHHHHCCTTTTTTBTHHHHHHHC..

SEQ PEQSSTAAASASATDTATPGAAGGATAAAAGSATSAPGDAARAASDNVAPADRSATP
 SEG ..xx.....
 1a06-

SEQ ATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQS
 SEG
 1a06-

SEQ SAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS
 SEG
 1a06-

Prosites for DKFZphfbr2_82e4.1

PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	51->54	PKC_PHOSPHO_SITE	PDOC00005
PS00005	91->94	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	118->121	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	264->267	PKC_PHOSPHO_SITE	PDOC00005
PS00005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS00005	454->457	PKC_PHOSPHO_SITE	PDOC00005
PS00005	467->470	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	118->122	CK2_PHOSPHO_SITE	PDOC00006
PS00006	248->252	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	336->340	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	467->471	CK2_PHOSPHO_SITE	PDOC00006
PS00007	456->464	TYR_PHOSPHO_SITE	PDOC00007
PS00007	127->136	TYR_PHOSPHO_SITE	PDOC00007
PS00008	260->266	MYRISTYL	PDOC00008
PS00008	321->327	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00009	59->63	AMIDATION	PDOC00009

Pfam for DKFZphfbr2_82e4.1

HMM_NAME	Eukaryotic protein kinase domain		
HMM	*YeigRiIGeGsFGtVYkCiWr.TGeIVAiKIkkrsms.....FlREIq		
Query	24	YDLGQVIKTEEFCEIFRAKDKTGKLTCKKFQKRDGRKVRKAANEIG	72
HMM	IMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEwe		
Query	73	ILKMVKHPNIIQLVDVFV-TRKEYFIFLELATGREVFDWILDQGYYSERD	121
HMM	IrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeN...gqIKicDFGLAR		
Query	122	TSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAK	171
HMM	qMnnYerMttfCGTPWY*		
Query	172	LEN--GLIKEPCGTPEY	186
HMM	*GepPFyd.....dnMemImrIiqrrfrpfWpnCSeElyDFMr		
Query	188	GNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSFYWDDISQAADLVT	236
HMM	wCWnyDPekRPTFrQILnHPWF*		
Query	237	RLMEVEQDQRITAEAAISHEWI	258

DKFZphfbr2_82g14

group: transmembrane protein

DKFZphfbr2_82g14 encodes a novel 208 amino acid proline-rich protein without similarity to known proteins.

The protein contains one transmembrane domain.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

unknown prolin rich protein

membrane regions: 1

Summary DKFZphfbr2_82g14 encodes a novel 208 amino acid protein.

unknown prolin rich protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 2059 bp

Poly A stretch at pos. 2049, polyadenylation signal at pos. 2024

```
1 AGAAGTGC GA CTGCCAGCTG CCGAGGCGTT CGGTCCTGCT GTTGCGGCCG
51 CTGCCCCCAGG GCTGCGGGGA CGCTCCCGGA GCCCTGCCTG TCCCCTGTCC
101 ATCCAGGCCA GCAGCTGAAG GAGCCTCACC TGCCTCCCTT CTCTGAGTAG
151 CACGGATTTC AGGAGAAGCA GCGAAGATGT CCAGCGAGCC TCCCCCTCCT
201 TATCCTGGGG GCCCCACAGC CCCACTTCTG GAAGAGAAAA GTGGAGCCCC
251 GCCCACCCCA GGCCGTTCCT CCCACAGCTGT GATGCAGCCC CCTCCAGGCA
301 TGCCACTGCC CCCTGCGGAC ATTGGCCCCC CACCCTATGA GCCGCCGGGT
351 CACCCAATGC CCCAGCCTGG CTTTCATCCCA CCACACATGA GTGCAGATGG
401 CACCTACATG CTTCCGGGTT TCTACCTTCC TCCAGGCCCC CACCCACCCA
451 TGGGCTACTA CCCCCAGGG CCCTACACGC CAGGGCCCTA CCCTGGCCCT
501 GGGGGGCCCA CAGCCACAGT CCTGGTCCCT TCAGGAGCTG CCACCACGGT
551 GACAGTGCTG CAGGGAGAGA TCTTTGAGGG AGCGCCTGTG CAGACGGTGT
601 GTCCCCACTG CCAGCAGGCC ATCGCCACCA AGATCTCCTA CGAGATTGGC
651 TTGATGAATT TCGTGCTGGG TTTCTTCTGT TGCTTCATGG GATGTGATCT
701 GGGCTGCTGC CTGATCCCTT GCCTCATCAA TGACTTCAAG GATGTGACGC
751 ACACATGCCC CAGCTGCAAA GCCTACATCT ACACGTACAA CGCCCTGTGC
801 TAACGGAGCT GGGACTCGGG ACTCCCCCGC CTGTCACTCT GGGCCCTGTG
851 GCTTTGTCTC CTGCGCTCAG TGGTCACTTT CCCGCTCCCA CTTGGGGCTG
901 GGAGCCGTGC CACCATCCCC TAGAAGTCCT GTCCTCTTCA CCCTGCCCTA
951 CCGAGCCCGC TGA CTCTTCTT GGCAAAAATT CTGTTGGGAT TTAAGGCCAA
1001 GGGTCAGTGG GTGGCAGGGG GCTGGCAATG AGCTTGTGTG TTGTTGGTCT
1051 GCTTGTGTGT TGTGATCGGG AAGATAAGCT GGGAGGGGTC TCCTGCTGGG
1101 GTCCTGATGC CTCTGTTTCC AAACAAGGTA CAGGTTCACT CCAGACTCTT
1151 TCCCCCTGGG ACCAACAGCA GCCAGAGCAG TTAGCCAGTT AGTCCCCAGG
1201 CCTGTGGCCA CAGGCGTTTC TGACCTGCTG GGCCGAGAAT GGGTAAGTTG
1251 TCTGGAGTCA GGTGGGCCCA CGTAGGACAG GGTCACAAAG CCTGGGTTTG
1301 TTTCTGGGTA CTTTGC GCCT CTGGGGTGCT AGAGGTGGGG CATGGTGGCT
1351 GGAAGTAAAA CTGCCAACTC TGGCCCTCAG AACTCTCAGG TATAGAAGCC
1401 CAGGATGTCT AATACCCTGT CCCAGTGCCC GAGAGCTGCC TGGTGTGAGG
1451 TAGAGAGGAC ACTGTACCTG GGTGAATGAT CAGACCCTGG TAGCTAAGAA
1501 GGAAC TTGTG CTTTGTGAGT AGTGTGCAGA CCCCCTTTCA GGCCATGCCT
1551 CTGTGAACCC TGTATTGCTG GGGCCGGAAG GAGCCCCTGA GCCTAGCCCC
1601 TTCCCCGTCTG CCCTGTGTCC TCACTGCGTG TGGGTATGAC CTCTGCCTGG
1651 TGCTGGTGT ATCCCAACTG GGCAAGAGAT GGCAGAGGGT CCCCCTGTGT
1701 GGTGCGCTTG GATGTGCAGA GCCTTCTCCA TGGATTTTCT TCCCTGTAAG
1751 TGCCCGGGCC CCCACCCAG CTGACAGGCT GTTGCTGTGC CTGCTCACAC
1801 CTGCTCCTGC AGGCACACTG GGCTAGGGAC GAGGAAGGAG CAGCCACAAG
1851 TGGTAGAACT GCCTTGGTGG ACACCAGCCT CGCCCTGTCT TTATTTCCTG
1901 AATGGTTTGT GAAC TTGCTC ACCTGGACCA CTGTATCCTG CCACTGTCTT
1951 TCCTGTCTCT GCACTGCCAC TGCTATGGCT CCTGTCACTG TGAATCGTGG
2001 CCCAGTCTCA GTTTGTAGTT TCTATTAA TTGGCCCTTT CACTCCCCCA
2051 AAAAAAAAAA
```

BLAST Results

Entry HS727347 from database EMBL:
 human STS WI-16589.
 Length = 275
 Plus Strand HSPs:
 Score = 1365 (204.8 bits), Expect = 3.0e-55, P = 3.0e-55
 Identities = 275/276 (99%), Positives = 275/276 (99%), Strand = Plus /
 Pl

Medline entries

No Medline entry

Peptide information for frame 3

```

1 MSSEPPPPYP GGPTAPLLEE KSGAPPTPGR SSPAVMQPPP GMPLPPADIG
51 PPPYEPPGHP MPQPGFIPPH MSADGTYMPP GFYPPPGPHF PMGYYPGPHY
101 TPGPYPGPGG HTATVLVPSG AATTVTVLQG EIFEGAPVQT VCPHCQQAIA
151 TKISYEIGLM NFVLGFFCCF MGCDLGCCLI PCLINDFKDV THTCPSCKAY
201 IYTYKRRLC

```

ORF from 177 bp to 800 bp; peptide length: 208
 Category: similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82g14, frame 3

PIR:S57447 HPBRII-7 protein - human, N = 1, Score = 206, P = 8.4e-16

PIR:A47655 spliceosome-associated protein SAP 62 - human, N = 1, Score
 = 198, P = 4.3e-15

>PIR:S57447 HPBRII-7 protein - human
 Length = 551

HSPs:

Score = 206 (30.9 bits), Expect = 8.4e-16, P = 8.4e-16
 Identities = 57/115 (49%), Positives = 62/115 (53%)

```

Query:      5 PPPPYPGGPTAPLLEEKSGAPPTPGRSSPAVMQPPPGMPLPPADIGPP-----PYEP--- 56
             PPPP+P G T P      G P PG P      PPPG LPP GPP P P
Sbjct:     226 PPPFPAGQTPP--RPPLGPPGPPGPPGP----PPPGQVLPPPLAGPPNRGDRPPPPVLF 279

```

```

Query:      57 PGHPMPQP--GFIPPHMSADGTYMPP-PGFYPPPGPHPPM-GYYPP-GPYTPGPYPGPGGH 111
             PG P QP G +PP      G P PG+ PPPGP PP G PP GP+ P P PGP G
Sbjct:     280 PGQPFQGPLGLPP-----GPPPPVPGYGPFGPPPPQQGPPPPGPFPPRP-PGPLGP 333

```

```

Query:     112 TATVLVP 118
             T+ P
Sbjct:     334 PLTLAPP 340

```

Score = 177 (26.6 bits), Expect = 1.1e-12, P = 1.1e-12
 Identities = 55/120 (45%), Positives = 61/120 (50%)

```

Query:      5 PPPPYPGGPTAP--LLEEKSGAPPTPG-RSSPAVM---QP---PPGMPLPPADIGPPPYE 55
             P PP P GP P +L      PP G R P V+ QP PP PLPP GPPP
Sbjct:     244 PGPPGPPGPPPPGQVLPPPLAGPPNRGDRPPPPVLFPGQPFQGPLGLPP---GPPP-P 299

```

```

Query:      56 PPGHPMPQPGFIPPHMSADGTYMPPPGFYPP--PGP-HPPMGYPPGPYTPGPYPG---PG 109
             PG+ P PG PP      G PPG +PP PGP PP+ PP P+ PGP PG P
Sbjct:     300 VPGYG-PPPGPPPPQQ---GPPPPPGPFPPRPPGLGPPPLTLAPP-PHLPGPPPGAPPPA 354

```

```

Query:     110 GHTATVLVP 118
             H      P
Sbjct:     355 PHVNPAFFP 363

```

Score = 168 (25.2 bits), Expect = 1.1e-11, P = 1.1e-11
 Identities = 47/118 (39%), Positives = 51/118 (43%)

```

Query:      5 PPPPYPG-GPTAPLLEEKSGAPPTPGRSSPAVMQP--PPGMPLPPADI-GPPFYEPGHP 60

```


DKFZphfbr2_82i17

group: signal transduction

DKFZphfbr2_82i17 encodes a novel 334 amino acid protein with similarity to the plasma membrane substrate for the cAMP-dependent protein kinase.

The novel protein is a transmembrane protein with strong similarity to the phospholemman protein, a membrane substrate for the cAMP-dependent protein kinase. It seems to serve as a chloride channel or as a chloride-channel regulator.

The new protein can find application in modulating/blocking cAMP-dependent protein kinase-dependent pathways.

similarity to plasma membrane substrate for cAMP-dependent protein kinase

complete cDNA, complete cds, EST hits
potential start at Bp 31 matches Kozak consensus PynNatgG
might be a SODIUM/POTASSIUM-TRANSPORTING ATPASE
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="11: 920_E_12; 786_(A,H)_11; (797,802)_(E,H)_7"

Insert length: 1647 bp

Poly A stretch at pos. 1637, polyadenylation signal at pos. 1615

```
1 AGTCTCGGAG GGGACCGGCT GTGCAGACGC CATGGAGTTG GTGCTGGTCT
51 TCCTCTGCAG CCTGCTGGCC CCCATGGTCC TGGCCAGTGC AGCTGAAAAG
101 GAGAAGGAAA TGGACCCCTT TCATTATGAT TACCAGACCC TGAGGATTGG
151 GGGACTGGTG TTCGCTGTGG TTCTCTTCTC GGTGGGATC CTCCTTATCC
201 TAAGTCCAG GTGCAAGTGC AGTTTCAATC AGAAGCCCCG GGCCCCAGGA
251 GATGAGGAAG CCCAGGTGGA GAACCTCATC ACCGCCAATG CAACAGAGCC
301 CCAGAAAGCA GAGAACTGAA GTGCAGCCAT CAGGTGGAAG CCTCTGGAAC
351 CTGAGGCGGC TGCTTGAACC TTGGATGCA AATGTCGATG CTTAAGAAAA
401 CCGGCCACTT CAGCAACAGC CCTTTCCCA GGAGAAGCCA AGAAGTTGTG
451 TGTCCCCAC CCTATCCCT CTAACACCAT TCCTCCACCT GATGATGCAA
501 CTAACACTTG CCTCCCCGCT GCAGCCTGTG GTCCCTGCCA CCTCCCGTGA
551 TGTGTGTGTG TGTGTGTGTG TGTGTGACTG TGTGTGTTTG CTAAGTGTGG
601 TCTTTGTGGC TACTTGTGTT TGGATGGTAT TGTGTTGTT AGTGAAGTGT
651 GGACTCGCTT TCCAGGCGAG GGGCTGAGCC ACACGGCCAT CTGCTCCTCC
701 CTGCCCCCGT GGCCCTCCAT CACCTTCTGC TCCTAGGAGG CTGCTTGTGG
751 CCCGAGACCA GCCCCCTCCC CTGATTTAGG GATGCGTAGG GTAAGAGCAC
801 GGGCAGTGGT CTCAGTCTGT CTGGGACCT GGAAGGTTT GCAGCACTTT
851 GTCATCATTC TTCATGGACT CCTTCACTC CTTAACAACA AACCTTGCTT
901 CCTTATCCCA CTGATCCCA GTCTGAAGGT CTCTTAGCAA CTGGAGATAC
951 AAAGCAAGGA GCTGGTGAGC CCAGCGTTGA CGTCAGGCG GCTATGCCCT
1001 TCCGTGGTTA ATTTCTTCCC AGGGGCTTCC ACGAGGAGTC CCCATCTGCC
1051 CCGCCCTTTC ACAGAGCGCC CGGGGATTCC AGGCCAGGG CTCTACTCT
1101 GCCCCTGGGG AATGTGTCCC CTGCATATCT TCTCAGCAAT AACTCCATGG
1151 GCTCTGGGAC CCTACCCCTT CCAACCTTCC CTGCTTCTGA GACTTCAATC
1201 TACAGCCCAG CTCATCCAGA TGCACTAC AGTCCCTGCA ATTGGGTCTC
1251 TGGCAGGCAA TAGTTGAAG ACTTCCTGTT CCGTTGGGGC CAGCACACCG
1301 GGATGGATGG AGGGAGAGCA GAGGCCTTTG CTTCTGTGCC TACGTCCCTT
1351 TAGATGGGCA GCAGAGGCAA CTCCCGCATC CTTGTCTGT CCTGTCAAGT
1401 GTCAGAGCGG TGAGCGAGGT GGGTTGGAGA CTCAGCAGGC TCCGTGCAGC
1451 CCTTGGGAAC AGTGAGAGGT TGAAGGTCAT AACGAGAGTG GGAAGTCAAC
1501 CCAGATCCCG CCCCTCCTGT CCTCTGTGTT CCCGCGGAAA CCAACCAAAAC
1551 CGTGCCTGTG GACCCATTGC TGTCTCTGT ATCGTGACCT ATCCTCAACA
1601 ACAACAGAAA AAAGGAATAA AATATCCTTT GTTCTCTAAA AAAAAAA
```

BLAST Results

Entry HS31455 from database EMBL:
human STS WI-2739.
Length = 103
Minus Strand HSPs:
Score = 487 (73.1 bits), Expect = 4.4e-14, P = 4.4e-14
Identities = 101/104 (97%), Positives = 101/104 (97%), Strand = Minus /
Plus
frame shift in primer binding site

Medline entries

91250422:

Purification and complete sequence determination of the major plasma membrane substrate for cAMP-dependent protein kinase and protein kinase C in myocardium.

95091702:

Protein kinase C and cyclic AMP-dependent protein kinase phosphorylate phospholemman, an insulin and adrenaline-regulated membrane phosphoprotein, at specific sites in the carboxy terminal domain.

95138184:

Mat-8, a novel phospholemman-like protein expressed in human breast tumors, induces a chloride conductance in *Xenopus* oocytes.

Peptide information for frame 2

1 MELVLVFLCS LLAPMVLASA AEKEKEMDPF HYDYQTLRIG GLVFAVVLFS
51 VGILLILSRR CKCSFNQKPR APGDEEAQVE NLITANATEP QKAEN

ORF from 32 bp to 316 bp; peptide length: 95
Category: strong similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82i17, frame 2

SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 196, P = 1.2e-15

TREMBL:AF091390_1 product: "phospholemman precursor"; *Mus musculus* phospholemman precursor, gene, complete cds., N = 1, Score = 187, P = 1.1e-14

PIR:A40533 cAMP-dependent protein kinase major membrane substrate precursor - dog, N = 1, Score = 189, P = 6.5e-15

SWISSPROT:PLM_RAT PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 185, P = 1.7e-14

>SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR.
Length = 92

HSPs:

Score = 196 (29.4 bits), Expect = 1.2e-15, P = 1.2e-15
Identities = 43/85 (50%), Positives = 56/85 (65%)

Query: 4 VLVFLCSLLAPMVLASAAEKEKEMDPFHYDYQTLRIGGLVFAVVLFSVGILLILSRRCKC 63
+LVF LL +AE KE DPF YDQ+L+IGGLV A +LF +GIL++LSRRC+C
Sbjct: 7 ILVFCVGLLT----MAKAESPKEHDPFTYDQSLQIGGLVIAGILFILGILVLSRRRC 62

Query: 64 SFNQKPRA--PGDEEAQVENLITANAT 88
FNQ+ R P +EE + I +T
Sbjct: 63 KFNQQQRTGEPDEEEGTFRSSIRRLST 89

Pedant information for DKFZphfbr2_82i17, frame 2

Report for DKFZphfbr2_82i17.2

[LENGTH] 95
[MW] 10542.37
[pI] 5.05
[HOMOL] SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR. 3e-15
[BLOCKS] BL01310

```

[EC]          3.6.1.37 Na+/K+-exchanging ATPase 6e-08
[PIRKW]       transmembrane protein 1e-09
[PIRKW]       hydrolase 6e-08
[PROSITE]     ATP1G1_PLM_MAT8      1
[PROSITE]     MYRISTYL             1
[PROSITE]     CK2_PHOSPHO_SITE     1
[PROSITE]     TYR_PHOSPHO_SITE     1
[PROSITE]     PKC_PHOSPHO_SITE     2
[PROSITE]     ASN_GLYCOSYLATION    1
[KW]          Alpha_Beta
[KW]          SIGNAL_PEPTIDE 19

```

```

SEQ  MELVLVFLCSLLAPMVLASAAEKEKEMDPFHYDYQTLRIGGLVFAVVLFSVGILLILSR
PRD  ccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  CKCSFNQKPRAPGDEEAQVENLITANATEPQKAEN
PRD  hhccccccccccccchhhhhhhhhhhcccccccccc

```

Prosite for DKFZphfbr2_82i17.2

PS00001	86->90	ASN_GLYCOSYLATION	PDOC00001
PS00005	36->39	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00007	25->33	TYR_PHOSPHO_SITE	PDOC00007
PS00008	41->47	MYRISTYL	PDOC00008
PS01310	28->42	ATP1G1_PLM_MAT8	PDOC01014

(No Pfam data available for DKFZphfbr2_82i17.2)

DKFZphfbr2_82i24

group: nucleic acid management

DKFZphfbr2_82i24 encodes a novel 547 amino acid protein with similarity to DEAD-box superfamily ATP-dependent helicases.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis.

The novel protein contains a DEAD-box an ATP/GTP-binding site motif A (P-loop, interacting with one of the phosphate groups of the nucleotide) and a leucine zipper. Mutations in the closely related *Drosophila* Hlc gene result in lethality in homozygotes. Therefore the new protein seems to be critical involved in RNA processing in eukaryotic cells.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to DEAD-box subfamily ATP-dependent helicase

complete cDNA, complete cds, EST hits
potential Start at Bp 9 matches Kozak consensus PyNNatgG,
[PFAM] Helicases conserved C-terminal domain
[PFAM] DEAD and DEAH box helicases

Sequenced by DKFZ

Locus: /map="720_A_3; 758_H_4; 772_E_3; 804_A_5; 175.5 cR from topFT of Chr7 linkage group"

Insert length: 1860 bp

Poly A stretch at pos. 1850, polyadenylation signal at pos. 1829

```
1 AGCAGCGCCA TGGAGGACTC TGAAGCACTG GGCTTCGAAC ACATGGGCCT
51 CGATCCCCGG CTCCTTCAGG CTGTCACCGA TCTGGGCTGG TCGCGACCTA
101 CGCTGATCCA GGAGAAGGCC ATCCCACTGG CCCTAGAAGG GAAGGACCTC
151 CTGGCTCGGG CCCGCACGGG CTCCGGGAAG ACGGCCGCTT ATGCTATTCC
201 GATGCTGCAG CTGTTGCTCC ATAGGAAGGC GACAGGTCCG GTGGTAGAAC
251 AGGCAGTGAG AGGCCTTGTT CTTGTTCTTA CCAAGGAGCT GGCACGGCAA
301 GCACAGTCCA TGATTCAGCA GCTGGCTACC TACTGTGCTC GGGATGTCCG
351 AGTGGCCAAT GTCTCAGCTG CTGAAGACTC AGTCTCTCAG AGAGCTGTGC
401 TGATGGAGAA GCCAGATGTG GTAGTAGGGA CCCCATCTCG CATATTAAGC
451 CACTTGCAGC AAGACAGCCT GAAACTTCGT GACTCCCTGG AGCTTTTGGT
501 GGTGGACGAA GCTGACCTTC TTTTTCCTT TGGCTTTGAA GAAGAGCTCA
551 AGAGTCTCCT CTGTCACCTG CCCCGGATT ACCAGGCTTT TCTCATGTCA
601 GCTACTTTTA ACGAGGACGT ACAAGCACTC AAGGAGCTGA TATTACATAA
651 CCCGGTTACC CTTAAGTTAC AGGAGTCCCA GCTGCCTGGG CCAGACCACT
701 TACAGCAGTT TCAGGTGGTC TGTGAGACTG AGGAAGACAA ATTCTCTCTG
751 CTGTATGCCC TGCTCAAGCT GTCATTGATT CGGGGCAAGT CTCTGCTCTT
801 TGTCAACACT CTAGAACGGA GTTACCGGCT ACGCCTGTTC TTGGAACAGT
851 TCAGCATCCC CACCTGTGTG CTCAATGGAG AGCTTCCACT GCGCTCCAGG
901 TGCCACATCA TCTCACAGTT CAACCAAGGC TTCTACGACT GTGTCATAGC
951 AACTGATGCT GAAGTCCTGG GGGCCCCAGT CAAGGGCAAG CGTCGGGGCC
1001 GAGGGCCCAA AGGGGACAA GGCCTCTGATC CGGAAGCAGG TGTGGCCCGG
1051 GGCATAGACT TCCACCATGT GTCTGCTGTG CTCAACTTTG ATCTTCCCCC
1101 AACCCCTGAG GCCTACATCC ATCGAGCTGG CAGGACAGCA CGCGCTAACA
1151 ACCCAGGCAT AGTCTTAACC TTTGTGCTTC CCACGGAGCA GTTCCACTTA
1201 GGCAAGATTG AGGAGCTTCT CAGTGGAGAG AACAGGGGCC CCATTCTGCT
1251 CCCTACCCAG TTCGGGATGG AGGAGATCGA GGGCTTCCGC TATCGCTGCA
1301 GGGATGCCAT GCGCTCAGTG ACTAAGCAGG CCATTCCGGG GSCAAGATTG
1351 AAGGAGATCA AGGAAGAGCT TCTGCATTCT GAGAAGCTTA AGACATACTT
1401 TGAAGACAAC CCTAGGGACC TCCAGCTGCT GCGGCATGAC CTACCTTTGC
1451 ACCCCGCAGT GGTGAAGCCC CACCTGGGCC ATGTTCTCTG CTACCTGGTT
1501 CCTCTGCTC TCCGTGGCCT GTTACGCCCT CACAAGAAGC GGAAGAAGCT
1551 GTCTTCTCTT TGTAGGAAGG CCAAGAGAGC AAAGTCCAG AACCCTATGC
1601 GCAGCTTCAA GCACAAAGGA AAGAAATTCA GACCCACAGC CAAGCCCTCC
1651 TGAGGTTGTT GGGCCTCTCT GGAGCTGAGC ACATTGTGGA GCACAGGCTT
1701 ACACCTCTCG TGGACAGGCG AGGCTCTGGT GCTTACTGCA CAGCCTGAAC
1751 AGACAGTTCT GGGGCCGGCA GTGCTGGGCC CTTAGTCTC TTGGCACTTC
1801 CAAGCTGGCA TCTTGCCCTC TGACAACAGA ATAAAAATTT TAGCTGCCCC
1851 AAAAAAAAAA
```

BLAST Results

Entry HSG05793 from database EMBL:
 human STS WI-6581.
 Length = 206
 Minus Strand HSPs:
 Score = 992 (148.8 bits), Expect = 6.0e-38, P = 6.0e-38
 Identities = 204/208 (98%), Positives = 204/208 (98%), Strand = Minus /
 P1

Entry AC004938 from database EMBL:
 Homo sapiens clone DJ0971C03; HTGS phase 1, 18 unordered pieces.
 Score = 1269, P = 6.5e-202, identities = 269/282
 12 exons Bp ~87920-93706 (matching 1-1497)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 10 bp to 1650 bp; peptide length: 547
 Category: strong similarity to known protein
 Classification: Nucleic acid management
 Prosite motifs: ATP_GTP_A (51-59)
 LEUCINE_ZIPPER (149-171)

```

1 MEDSEALGFE HMGLDPRLLQ AVTDLGWSRP TLIQEKAIP L ALEGKDLLAR
51 ARTGSGKTAA YAIPLQLLL HRKATGPVVE QAVRGLVLP TKELARQAQS
101 MIQLATYCA RDVRVANVSA AEDSVSQRV LMEKPDVVVG TPSRILSHLQ
151 QDSLKLRLSL ELLVVDEADL LFSFGFEEEL KSLCHLPRI YQAFMSATF
201 NEDVQALKEL ILHNPVTLKL QESQLPGPDQ LQQFQVVCET EEDKFLLLYA
251 LLKLSLIRGK SLLFVNTLER SYRLRLFLEQ FSIPTCVLNG ELPLRSRCHI
301 ISQFNQGFYD CVIATDAEVL GAPVKGKRRG RGPKGDKASD PEAGVARGID
351 FHHVAVLNF DLPPTPEAYI HRAGRTARAN NPGIVLTFVL PTEQFHLGKI
401 EELLSGENRG PILLPYQFRM EEIEGFYRC RDAMRSVTQK AIREARLKEI
451 KEELLHSEKL KTYFEDNPRD LQLLRHDLPL HPAVVKPHLG HVPDYLVPFA
501 LRGLVRPHKK RKKLSSSSCRK AKRAKSQNPL RSFKHKGKKF RPTAKPS

```

BLASTP hits

No BLASTP hits available
 Alert BLASTP hits for DKFZphfbr2_82i24, frame 1

TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds., N = 1, Score = 1230, P = 3.2e-125

TREMBL:SPCC1494_6 gene: "SPCC1494.06c"; product: "atp dependent helicase"; S.pombe chromosome II cosmid c1494., N = 2, Score = 753, P = 2.5e-113

PIR:S51412 hypothetical protein YLR276c - yeast (Saccharomyces cerevisiae), N = 2, Score = 711, P = 8.2e-117

TREMBL:AF025451_2 gene: "C24H12.4"; Caenorhabditis elegans cosmid C24H12., N = 2, Score = 564, P = 2.7e-99

>TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds.
 Length = 560

HSPs:

Score = 1230 (184.5 bits), Expect = 3.2e-125, P = 3.2e-125
 Identities = 251/497 (50%), Positives = 344/497 (69%)

Query: 9 FEHMLDPRLLQAVTDLGWSRPTLIQEKAIPALEGKOLLARARTGSGKTAAYAIPMLQL 68
 F + LD R+L+AV LGW +PTLIQ AIPL LEGKD++ RARTGSGKTA YA+P++Q
 Sbjct: 11 FHELELDQRIKAVAGLQGWQQPTLIQSTAIPLLEGGKDVVVRARTGSGKTATYALPLIQ 70

Query: 69 LLHRKATGPVVEQAVRGLVLVPTKELARQAQSMIQQLATYCARDVRVANVS-AAEDSVSQ 127
 +L+ K EQ V +VL PTKEL RQ++ +I+QL C + VRVA+++ ++ D+V+Q
 Sbjct: 71 ILNSKLNAS--EQYVSAVVLAPTKELCRQSRKVIEQLVESCGKVVVRADIADSSNDTVTQ 128

Query: 128 RAVLMEKPDVVVGTSPSRILSHLQQDSLKLRLDSLELLVVDEADLLFSFGFEEELKSLLCHL 187
 R L E PD+VV TP+ +L++ + S+ +E LVVDEADL+F++G+E++ K L+ HL
 Sbjct: 129 RHALESPIDIVATPANLLAYAEAGSVVDLKHVETLVVDEADLVFAYGYEKDFKRLIKHL 188

Query: 188 PRIYQAFILMSATFNEDVQALKELILHNPVTILKQESQLPGPDQLQQFQVVCETEEDKFL 247
 P IYQA L+SAT +DV +K L L+NPVTILK+L +L DQL +++ E E DK +
 Sbjct: 189 PPIYQAVLVSATLTDVVRMKGLCLNPNVTILKLEPELVPODQLSHQRILAE-ENDKPAI 247

Query: 248 LYALLKLSLIRGKSLLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHIISQFNQG 307
 LYALLKL LIRGKS++FVN+++R Y++RLFLEQF I CVLN ELP R H ISQFN+G
 Sbjct: 248 LYALLKRLIRGKSIIFVNSIDRCYKVRFLFLEQFGIRACVLNSEL PANIRIHTISQFNKG 307

Query: 308 FYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPE 367
 YD +IA+D + P G + K ++ D E+ +RGIDF V+ V+NFD P
 Sbjct: 308 TYDIIIASDEHHMEKP--GGKSATNRKSPRSGDMESSASRGIDFQCVNNVINFDPRDVT 365

Query: 368 AYIHRAGRTARANNPGIVLTFVLPTEQFHLGKIEELL----SGENRGPIILLPYQFRMEEI 423
 +YIHRAGRTAR NN G VL+V E +E+ L + + I+ YQF+MEE+
 Sbjct: 366 SYIHRAGRTARGNNKGSVLSFVSMKESKVNDSEVKKLDCSFAAQEGEQIKNYQFKMEEV 425

Query: 424 EGFYRCRDAMRSVTKQAIAREARKKEILLHSEKLKTYFEDNPRDLQLLRHDLPLHPA 483
 E FRYR +D R+ T+ A+ + R++EIK E+L+ EKLK +FE+N RDLQ LRHD PL
 Sbjct: 426 ESFRYRAQDCWRAATRVAVHDTRIIEIKIEILNCEKKAFFENKRDLQALRHDKPLRAI 485

Query: 484 VVKPHLGHVDPDYLVPPALRGLV 505
 V+ HL +P+Y+VP AL+ +V
 Sbjct: 486 KVQSHLSDMPEYIVPKALKRVV 507

Pedant information for DKFZphfbr2_82i24, frame 1

Report for DKFZphfbr2_82i24.1

[LENGTH] 547
 [MW] 61589.88
 [PI] 9.34
 [HOMOL] TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster
 tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent
 bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst),
 and la costa (lcs) genes, complete cds. 1e-121
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR276c] 1e-109
 [FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA]
 2e-42
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YLL008w] 8e-40
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 8e-40
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YLL008w] 8e-40
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S.
 cerevisiae, YKR059w] 3e-39
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YKR059w] 3e-39
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 3e-35
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 3e-29
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 4e-29
 [FUNCAT] 1 genome replication, transcription, recombination and repair [H.
 influenzae, HI0892] 1e-27
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-27
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 4e-21
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 1e-05
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-34
 [PIRKW] RNA binding 7e-41
 [PIRKW] DEAD box 2e-38
 [PIRKW] transmembrane protein 9e-20
 [PIRKW] DNA binding 8e-23
 [PIRKW] ATP 1e-107
 [PIRKW] purine nucleotide binding 2e-38
 [PIRKW] P-loop 1e-107
 [PIRKW] hydrolase 2e-35
 [PIRKW] protein biosynthesis 2e-38
 [PIRKW] ATP binding 7e-43

```

[SUPFAM]      WW repeat homology 1e-26
[SUPFAM]      DEAD/H box helicase homology 1e-107
[SUPFAM]      unassigned DEAD/H box helicases 1e-107
[SUPFAM]      ATP-dependent RNA helicase DBP1 3e-31
[SUPFAM]      ATP-dependent RNA helicase DHH1 2e-35
[SUPFAM]      translation initiation factor eIF-4A 2e-38
[SUPFAM]      tobacco ATP-dependent RNA helicase DB10 1e-26
[PROSITE]     ATP_GTP_A      1
[PROSITE]     LEUCINE_ZIPPER 1
[PFAM]        Helicases conserved C-terminal domain
[PFAM]        DEAD and DEAH box helicases
[KW]          Alpha_Beta
[KW]          LOW COMPLEXITY      9.87 %

```

```

SEQ      MEDSEALGFHEHMGDPRLLQAVTDLGWSRPTLIQEKAIPALALEGKDLLARARTGSGKTA
SEG
PRD      cccccccccccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccc

```

SEQ YAI PMLQLLLHRKATGPVVEQAVRGLVLVPTKELARQAQSMIQLATYCARDVRVANVSA
 SEG
 PRD ehhhhhhhhhhhhccccccccceeeeeccchhhhhhhhhhhhhhhhhhhhhccccceeeec

SEQ AEDSVSQRAVLMEKPDVVVGTPSRILSHLQDLSKLRLDSLELLVVDEADLLFSFGFEEEL
 SEGXXXXXXXXXXXXX.....
 PRD cscchhhhhhhhhccceeeccscchhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhcchhhh

SEQ KSL LCH LPR IYQ AFLMSATFNEDVQALKE LILHNPVTLKLQESQLPGPDQLQFQVVCET
 SEG
 PRD hhhhhhhccchhhhhhhhhhhhhccchhhhhhhhhhhhhccceeeeeeccccccchhhhhhhhhhhhhhh

```

SEQ      EEDKFLLYALLKLSLIRGKSLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHI
SEG      .....xxxxxxxxxxxxx.....
PRD      hhhhhhhhhhhhhhhhhccceeeeeehhhhhhhhhhhhhhhccceeccccchhhhhhhh

```

```

SEQ      ISQFNQGFYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNF
SEG      .....xxxxxxxxxxxxx.....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

SEQ DLPPTPEAYIHRAGRTARANNPGIVLTFVLPTREQFHLGKIEELLSGENRGPILLPYQFRM
 SEG:
 PRD cccccceeeecccccccccccccceeeecchhhhhhhhhhhhhhhcccccchhh

```
SEQ      EEIEGFRYRCRDAMRSVTKQAIREARLKEIKEELHSEKLKTYFEDNPRDLQLLRHDLPL  
SEG      .....  
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhccc
```

```
SEQ      HPAVVKPHLGHPDYLVPALRGLVRPHKKRKKLSSSCRKAKRAKSQNPLRSFKHKGKKF
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccccccccccceccccccccccccccccccccchhhhhcccccccccccccccccc
```

```

SEQ      RPTAKPS
SEG      . . . . .
PRD      CCCCCC

```

Prosites for DKFZphfbr2 82i24.1

PS00017	51->59	ATP_GTP_A	PDOC00017
PS00029	149->171	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphfbr2 82i24.1

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLpPWILRnIyeMGFEkPTPIQQaPiILeGRDVMAcAQTGSGKTAAF		
Query	13	GLP+P +L +++++G+++PT IQ++AIP++LEG+D++A+A TGSgKTAa+ GLDPRLLQAVTDLGWSRPTLIQEKAIPLALEGKDLLARARTGSGKTAAY	61
HMM	LIPMLQHIDwdP...WpqpQDPrALILAPtRELAMQIQEEcrkFGkHmn		
Query	62	+IPMLQ +++ + + + + + +R+L+L+PT ELA+Q + + + + + ++ AIPMLQLLLHRKATGPVVEQA-VRGLVLVPTKELARQAQSMIQLATYCA	110
HMM	g.IRImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIERgtldLDr.		
Query	111	+R++ + + Q +L+++P ++V++TP R++ H+++ +L+L++ RDVRVANVSAAEDSVSQRAVLMEKP-DVVVGTPSRILSHLQQDSLKLKRDs	159
HMM	IeMLVMDEADRMldMGFIQIRrImrqiPmpwNRQTMMFSATMPdeIqEL		
Query	111	+E LV DEAD +++ GF++++ ++ +P+ + O + SAT+ +++O L	

Query 160 LELLVVDEADLLFSFGFEEELKSLCHLP--RIYQAFILMSATFNEDVQAL 207
HMM ARrFMRNPiRInIdMdElTtnEnIkQwYiyVerEMWKfdCLcrLie*
+ +++NP+ + + +++L + ++Q+ +++E E++KF +L+ L++
Query 208 KELILHNPVTLKLQESQLPGPDQLQQFQVVCETEEDKFLLLYALLK 253

HMM_NAME Helicases conserved C-terminal domain

HMM *EilleeWLknIGIrvmYIHGdMpQeERdeIMddFNnGEynVLicTDV...
+L+ +L++ I++++ G +P + R I+ +FN+G Y++ I+TD+
Query 272 YRLRLFLEQFSIPTCVLNGELPLRSRCHIISQFNQGFYDCVIATDAEVL 320
HMMggRGIDIPdVNHVINYDMPWNPEqYI
+RGID+ V+ V N+D+P +PE YI
Query 321 GAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPEAYI 370
HMM QRIGRTgRIG*
+R+GRT+R++
Query 371 HRAGRTARAN 380

DKFZphfbr2_82m16

group: brain derived

DKFZphfbr2_82m16 encodes a novel 289 amino acid protein with very weak similarity to A.thaliana F28A23.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F28A23.140

complete cDNA, complete cds, few EST hits
many ATGs in front of the ORF
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="4"

Insert length: 2715 bp

Poly A stretch at pos. 2705, polyadenylation signal at pos. 2687

```
1 AGAGGAGGGG AGAGGACTGG GGAGCCGAGC CAGAGCCGGG CTGCCTGCCA
51 CCCGGCTGCT CGTCCGCTAG CTGGGGAGGA GCGCTCCACC CGCAACTGAC
101 AAAGGATGGG AGAATGCCCG CGCCCCGGGA TGCCGGCCGC ACGCAGCCTG
151 GCGGCCGCCT GAGCTACTTC ACCCTCCGCC GGTAAGTGAC TGCAAACATC
201 ATTCAATCAA TCAGCCTCAC TGGGAGCCCC TTCTCTCCGG CTGGTAGTCC
251 TGGGCGGCTT GTCCCTGATC CCGAGCGGGG CTGGGCACAG CATCAGCCCT
301 GGAGGGCAGG CAGCAGGTGC CTTTGCCTGG TGGGTCCACT GGGGAGCGTG
351 GCTGGGGTTC GCGGCGGGTG CTGCCACCCA ACCTGCGGGC GCGGGGCTCG
401 CCCAGTAGGC GCCTCTCTGG TGAGAGGAGG CGGCTCCAGC CCGCATCCTG
451 GGGTAGTTGC TACTATTGGC CCCAGCGGCC CGCTCTGCGC GCGCGCCGTT
501 TCTGGCGGAT CCCAGTGCGC GCGCGCGCTG TTTACACCGG CGTGGTACTA
551 GTCACGGAGC CGCACCCCTC GGAAGCGCGG GAGTCGATGA CAGCCACTTC
601 ACAGGGTCAC GCGCTCCTAG TGTGGGCTTG AAGGGGACGG GGACCGATTA
651 CCAAAGGAGA GCGCTGAGTA CGGAAGACAC AGGGCAGCCT TTGTCTTGGG
701 TTTAGCGCTG ATGCGCTCAA CCCTGAGTCG GGTTCCTGAC AACTGTTGTG
751 TCCGATTTTC GTTCCCTGCA ACCGCCCTCC TGGGCGAGAG ATGTCATTGT
801 GTTCTGCGCG CCAGCGGGAC TGAGAGCTGG GACTTAAGAC GCCAGGAGGG
851 TCCTGCGCTC ACGGGAAATG TACCCCAAAA GAACTCTGAG AGAATATACT
901 CAACTGTCCT GCTGTGATTA ACAAAGACTG CTGTATTTTA ATTTCAAGAA
951 TTGAAAAGGG ATAGGAGGAA GGGGAAAATG CTGGGCTGGT GTGAAGCGAT
1001 AGCCCCGTAAC CCTCACAGAA TTCCAAACAA CACGCGAACA CCGGAGATCT
1051 CAGGGGATTT GGCTGACGCC TCACAAACCT CCACATTGAA TGAAAAATCC
1101 CCAGGGCGAT CTGCAAGTCG ATCAAGTAAC ATTTCAAAG CAAGCAGCCC
1151 AACAAACAGG ACAGCTCCCA GGAGCCAGTC AAGGTTGTCT GTCTGTCCAT
1201 CCACTCAGGA CATCTGCAGA ATCTGTCACT GCGAAGGGGA TGAAGAGAGC
1251 CCCCTCATCA CACCCTGTCT CTGCACTGGG ACACTGCGCT TTGTCCACCA
1301 GTCCCTGCCTC CACCACTGGA TAAAGAGCTC AGATACACGC TGCTGTGAGC
1351 TCTGCAAGTA TGAATTCATA ATGGAGACCA AGCTCAAACC CCTCCGGAAG
1401 TGGGAGAAAC TACAGATGAC CACAAGTGAA AGGAGGAAAA TATTCTGCTC
1451 TGTCACATTC CACGTAATCG CGATCACCTG TGTGGTTTGG TCTTTGTATG
1501 TATTGATAGA CCGGACAGCG GAGGAAATCA AGCAAGGCAA TGACAATGGT
1551 GTCCTTGAAT GGCCATTTTG GACAAAACCT GTTGTGGTAG CCATTGGCTT
1601 CACAGGAGGT CTTGTCTTCA TGTACGTACA GTGTAAAGTC TATGTTCACT
1651 TGTGGCGCAG GCTGAAGGCC TACAACCGTG TGATCTTTGT ACAAATTTGC
1701 CCAGACACTG CCAAAAAACT GGAGAAGAAC TTCTCATGTA ATGTAAACAC
1751 AGACATCAAA GATGCTGTGG TAGTGCTGTG ACCACAAACA GGTGCAAAAT
1801 CACTGCCATC TGCAGAGGGT GGGCCCCCTG AAGTTGTATC AGTCTGATGG
1851 AACCTGTTGG GAGTTTCTTC ACCGAAGAAT ATCTTTCTAG CCCTCAGCCA
1901 CTACAAATGA CAGAAGTGAC CTTGAATTAT TTAATCCCTT CAGCTCCTCC
1951 TTTCTCCTAC TGACACATTT TTCCTGACTT TGTTCAAAGA GGAAGGAGA
2001 AAAACAAACA AACAGACCAA ATGCCCAGGA GCCCATGAAG TAATAGCGTA
2051 AAGTAAAGTA TGATATGGAA ATGTGAAGTT TGCAAGAGAA TGATTTCCAA
2101 GACAATTAAG AACTACTGGG GCAATGAATG CTTTATAGGCA GTAATCAAAG
2151 ATTAATGGGA CCCATGATAC TCTTCTTAC AGTAACAGGG GAAAAGTTCA
2201 AGAATACAGA CTTGAATTGC GATGTGTATT ACTTCTAGGG CCTTGTAATG
2251 TTAACCTGTCT CATCTGGAAA TAATAACTAA CATATTGGT TTTAAGCCTG
2301 AAATTGTCTG CATTATCCCT AAGTCACATT GGAAGTGAAC TTGGAGGATG
2351 CATATTTTGA TATGCTTTGA CAGCTAACAG ATTTGTATGG TTTAGTGGAG
2401 TCTGGTTATT TTGACAGATG CATGTTTTT TTAATAGAT GCAATATACA
2451 TTTGAAGACA TTGATATTTG GAATTAATTA TGTGTTGTTA AGTCACGCAA
2501 AAGATTTTCA GAAATGTTT GGATATAATT AGCTCTGTTA AATACCCACA
2551 GAACTGTTAT CAGGTCTTAT ATTTATTTTC ATCTGGTTCC TCTAATACAG
```


2601 TGCTGTCCAA TAGAAACACA ACAGCCACAA ATGCAGGCCA CAGATGCAAA
 2651 TATTTAACTT CCCAGTAGCC CTATTTTAAA AAGTAAAAAT AAATGTTTGT
 2701 TTGTTAAAAA AAAAA

BLAST Results

Entry G37457 from database EMBLNEW:
 SHGC-57357 Human Homo sapiens STS genomic.
 Length = 458
 Plus Strand HSPs:
 Score = 2116 (317.5 bits), Expect = 4.3e-91, P = 4.3e-91
 Identities = 444/456 (97%)

Medline entries

No Medline entry

Peptide information for frame 3

1 MLGWCEAIAR NPHRIPNNTR TPEISGDLAD ASQTSTLNEK SPGRSASRSS
 51 NISKASSPTT GTAPRSQSRL SVCSTQDIC RICHCEGDEE SPLITPCRCT
 101 GTLRFVHQSC LHWIKSSDT RCCELCKYDF IMETKLKPLR KWEKLQMTTS
 151 ERRKIFCSVT FHVIAITCVV WSLYVLIDRT AEEIKQNDN GVLEWPFWTK
 201 LVVVAIGFTG GLVFMVQCK VYVQLWRLK AYNRVIFVQN CPDTAKKLEK
 251 NFSCNVNTDI KDAVVVPVQ TGANSLPSAE GGPPEVSV

ORF from 978 bp to 1844 bp; peptide length: 289
 Category: similarity to unknown protein

BLASTP hits

Entry AB011169_1 from database TREMBL:
 gene: "KIAA0597"; product: "KIAA0597 protein"; Homo sapiens mRNA for
 KIAA0597 protein, partial cds.
 Score = 188, P = 6.0e-12, identities = 30/54, positives = 38/54

Entry SPBC14F5_7 from database TREMBL:
 gene: "SPBC14F5.07"; product: "hypothetical protein"; S.pombe
 chromosome II cosmid c14F5.
 Score = 185, P = 1.9e-11, identities = 29/53, positives = 38/53

Entry CEY57A10B_1 from database TREMBL:
 gene: "Y57A10B.1"; Caenorhabditis elegans cosmid Y57A10B
 Score = 171, P = 2.6e-10, identities = 40/107, positives = 58/107

Alert BLASTP hits for DKFZphfbr2_82m16, frame 3

TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII
 project), N = 1, Score = 198, P = 3.4e-13

>TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII project)
 Length = 1,051

HSPs:

Score = 198 (29.7 bits), Expect = 3.4e-13, P = 3.4e-13
 Identities = 38/103 (36%), Positives = 61/103 (59%)

Query: 28 LADASQTSTLNEKSPGRSASRS-SNISKASSPTTGTAPRSQSRLSVCSTQDICRICHCE 86
 +++ S +S+ + SP +++ SN+ A S TG+ +D+CRIC
 Sbjct: 20 VSEPSVSSSSSSSPNQASPNPFSNMDPAVSTATGSRVYDDDE----DEEDVCRICRNP 74

Query: 87 GDEESPLITPCRCTGTLRFVHQSC LHWIKSSDTRCCCELCKYDF 130
 GD ++PL PC C+G+++FVHQ CL QW+ S+ R CE+CK+ F
 Sbjct: 75 GDADNPLRYPACSGSIKFEVHQDCLLQWLHNSNARQCEVCKHPF 118

[illegible]

(No Pfam data available for DKFZphfbr2 82m16.3)

DKFZphfbr2_82m6

group: signal transduction

DKFZphfbr2_82m6.3 encodes a novel 654 amino acid protein with similarity to murine sphingosine kinase.

Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellularly, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependend on SPP. Extracellularly, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1.

The new protein can find application in modulating/blocking the shingosine kinase intracellular signal transmission pathway.

strong similarity to mouse "sphingosine kinase"

complete cDNA, complete cds, EST hits,
YLR260w/YOR171c Lcb5p/Lcb4p = long chain base kinases,
involved in biosynthesis of sphingolipids

Sequenced by DKFZ

Locus: unknown

Insert length: 2875 bp

Poly A stretch at pos. 2865, polyadenylation signal at pos. 2838

```
1 AGTGTGGAG GTGAGGAGG GGGGCTGGCA GGGCTAGTCG GGGCATCTGG
51 AAATTTCCGA CCCACGCTT CGGGCGTTTC CTTATCAGGT TCACCGCTCC
101 CTGATCTCGC GCTGCACCTT GTAGGCGCAG CCGCTGCTTG GGAAGTCTTA
151 CTTAAGAGCT GAAGGTCAGG CCAGGACAGT GAGACCTGAC TCCTTGCTCC
201 TACCAGCCTA CTATGGCTTA AGACCCAGGG CCAGGGTCCC GTTGATGTAA
251 CAGAGCAGAG GACCAGCAGA TGAATGGACA CCTTGAAGCA GAGGAGCAGC
301 AGGACCAGAG GCCAGACCAG GAGCTGACCG GGAGCTGGGG CCACGGGCCT
351 AGGAGCACCC TGGTCAGGGC TAAGGCCATG GCCCCGCCCC CACCGCCACT
401 GGGTGCAGC AGCTCGCTCC TCCATGGCGA GTTTGGCTCC TACCCAGCCC
451 GAGGCCACAG CTTTGCCCTC ACCCTTACAT CGCAGGCCCT GCACATACAG
501 CGGCTGCGCC CCAAACCTGA AGCCAGGCCC CGGGGTGGCC TGGTCCCGTT
551 GGGCAGGCTC TCAGGCTGCT GCACCCCTGC AGCCCGCAGC CCCTCAGACT
601 CAGCGGCCTA CTTCTGCATC TACACCTACC CTCGGGGCCG GCGCGGGGCC
651 CGGCGCAGAG CCACTCGCAC CTTCCGGGCA GATGGGGCCG CCACCTACGA
701 AGAGAACCGT GCCGAGGCCC AGCGCTGGGC CACTGCCCTC ACCTGTCTGC
751 TCCGAGGACT GCCACTGCCC GGGGATGGGG AGATCACCCC TGACCTGCTA
801 CCTCGGCCCG CCCGGTTGCT TCTATTGGTC AATCCCTTTG GGGGTCGGGG
851 CTGGGCTGCG CAGTGGTGTA AGAACCACGT GCTTCCCATG ATCTCTGAAG
901 CTGGGCTGTC CTTCAACCTC ATCCAGACAG AACGACAGAA CCACGCCCGG
951 GAGCTGGTCC AGGGGCTGAG CCTGAGTGAG TGGGATGGCA TCGTCACGGT
1001 CTCGGGAGAC GGGCTGCTCC ATGAGGTGCT GAACGGGCTC CTAGATCGCC
1051 CTGACTGGGA GGAAGCTGTG AAGATGCCTG TGGGCATCCT CCCCTGCGGC
1101 TCGGGCAACG CGCTGGCCGG AGCAGTGAAC CAGCACGGGG GATTTGAGCC
1151 AGCCCTGGGC CTCGACCTGT TGCTCAACTG CTCACTGTTG CTGTGCCGGG
1201 GTGGTGGCCA CCCACTGGAC CTGCTCTCCG TGACGCTGGC CTCGGGCTCC
1251 CGCTGTTTCT CTTCTCTGTC TGTGGCTTGG GGCTTCGTGT CAGATGTGGA
1301 TATCCAGAGC GAGCGCTTCA GGGCCTTGGG CAGTGCCCGC TTCACACTGG
1351 GCACGGTGCT GGGCCTCGCC AACTGACACA CCTACCGCGG ACGCCTCTCC
1401 TACCTCCCGC CCACTGTGGA ACCTGCCTCG CCCACCCCTG CCCATAGCCT
1451 GCCTCGTGCC AAGTCGGAGC TGACCCCTAAC CCCAGACCCA GCCCCGCCCA
1501 TGGCCCACTC ACCCCTGCAT CGTTCTGTGT CTGACCTGCC TCTTCCCTCG
1551 CCCAGCCTG CCTTGGCCTC TCCTGGCTCG CCAGAACCCC TGCCCATCCT
1601 CTCCTCAAC GGTGGGGGCC CAGAGCTGGC TGGGGACTGG GGTGGGGCTG
1651 GGGATGCTCC GCTGTCCCGC GACCCACTGC TGTCTTACC TCCTGGCTCT
1701 CCCAAGCGAC CTCTACACTC ACCCGTCTCC GAAGGGGCCC CCGTAATTCC
1751 CCCATCTCTT GGGCTCCAC TTCCCACCCC TGATGCCCGG GTAGGGGCTG
1801 CCACCTGCGG CCCGCCGAC CACCTGCTGC CTCGCTAGG CACCCCGCTG
1851 CCCCCAGACT GGGTGACGCT GGAGGGGGAC TTTGTGCTCA TGTGGCCAT
1901 CTCGCCGAGC CACCTAGGCG CTGACCTGGT GGCAGCTCCG CATGCGCGCT
1951 TCAGACGACG CTTGGTGAC CTGTGCTGGG TGGTAGCGG CATCTCGCGG
2001 GCTGCGCTGT TGGCGCTTTT CTTGGCCATG GAGCGTGGTA GCCACTCAG
2051 CTGGGCTGT GCTACGCGCG GCGCCGTGCC TTCCGCTAG
2101 AGCCGCTCAC ACCACGCGGC GTGCTACAG TGGACGGGGA GCAGGTGGAG
2151 TATGGGCCCG TACAGGCACA GATGCACCT GGCATCGGTA CACTGCTCAC
2201 TGGGCTCTCT GGCTGCCCGG GCGGGGAGCC CTGAACTAA ACAAGCTTGG
2251 TACCCGCGCG GGGCGGGGCC TACATTCCAA TGGGGCGGAG CCTGAGCTAG
2301 GGGGTGTGGC CTGGCTGCTA GAGTTGTGGT GGCAGGGGCC CTGGCCCCGT
```

```

2351 CTCAGGATTG CGCTCGCTTT CATGGGACCA GACGTGATGC TGAAGGTGG
2401 GCGTCGTAC GGTAAAGAG AAATGGGCTC GTCCCGAGGG TAGTGCCTGA
2451 TCAATAGGG CGGGCCCTGG CGTCTGATCT GGGGCCGCC TTACGGGCA
2501 GGGCTCAGTC CTGACGCTTG CCACCTGCTC CTACCCGCC AGGATGGCTG
2551 AGGGCGGAGT CTATTTTACG CGTCGCCAA TGACAGGACC TGAATGTAC
2601 TGGCTGGGGT AGGCCTCAGT GAGTCGGCG GTCAGGGCCC GCAGCCTCGC
2651 CCCATCCACT CCGGTGCCCTC CATTAGCTG GCCAATCAGC CCAGGAGGGG
2701 CAGGTTCCCC GGGGCCGGCG CTAGGATTG CACTAATGTT CCTCTCCCG
2751 CGGGTGGGGG CGGGGAAATT CATATCCCT GTTCGTCTCA TGCCTCCT
2801 CCGTCCCCAA TCTAAAAAGC AATTGAAAG GTCTATGCAA TAAAGGCAGT
2851 CGTTCATTC CTCTCAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

99045661:

Tumor necrosis factor-alpha induces adhesion molecule expression through the sphingosine kinase pathway.

98395082:

Molecular cloning and functional characterization of murine sphingosine kinase.

98241633:

Purification and characterization of rat kidney sphingosine kinase.

99178622:

Sphingosine 1-phosphate: a prototype of a new class of second messengers.

Peptide information for frame 3

```

1 MNHLEAEEQ QDQRPDQELT GSWGHGPRST LVRKAMAPP PPPLAASTSL
51 LHGEFGSYPA RGRPFALTLT SQALHIQRLR PKPEARPRGG LVPLAEVSGC
101 CTLRSRSPSD SAAYFCIYTY PRGRRGARRR ATRTFRADGA ATYEENRAEA
151 QRWATALTCL LRGLPLPGDG EITPDLLPRP PRLLLVNPF GGRGLAWQWC
201 KNHVLPmise AGLSFNLIQT ERQNHARELV QGLSLSEWDG IVTVSGDGLL
251 HEVLNGLLDR PDWEEAVKMP VGILPCGSGN ALAGAVNQHG GFEPALGLDL
301 LLNCSLLLCR GGGHPLDLS VTLASGSRCF SFLSVWGFV SDVDIQSERF
351 RALGSAREFTL GTVLGLATLH TYRGRLSYLP ATVEPASPTP AHSPLPRAKSE
401 LTLTPDPAPP MAHSPLHRV SDLPLPLPQP ALASPGSPEP LPILSLNGGG
451 PELAGDWGGA GDAPLSPDPL LSSPPGSPKA ALHSPVSEGA PVIPPSSGLP
501 LPTPDARVGA STCGPPDHLL PPLGTPLPPD WVTLEGDFVL MLAISSPHLG
551 ADLVAAPHAR FDDGLVHLCW VRSGISRAAL LRLFLAMERG SHFSLGCPQL
601 GYAAARAFRL EPLTPRGVLT VDGEQVEYGP LQAMHPGIG TLTGPPGCP
651 GREP

```

ORF from 270 bp to 2231 bp; peptide length: 654
Category: similarity to known protein

BLASTP hits

Entry SPAC4A8.7 from database TREMBL:

gene: "SPAC4A8.07c"; product: "hypothetical protein"; S.pombe chromosome I cosmid c4A8.

Score = 301, P = 7.9e-32, identities = 68/190, positives = 109/190

Entry CEC34C6.3 from database TREMBLNEW:

product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6

>TREMBL:CEC34C6_3 product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6

Score = 273, P = 9.0e-29, identities = 78/265, positives = 142/265

Entry S67059 from database PIR:

hypothetical protein YOR171c - yeast (Saccharomyces cerevisiae)

>TREMBL:SC55021_9 gene: "O3615"; product: "O3615p"; Saccharomyces cerevisiae cosmid pUA1258 from chromosome 15R. >TREMBL:SCYOR170W_2 S.cerevisiae chromosome XV reading frame ORF YOR170W

Score = 253, P = 2.0e-25, identities = 70/234, positives = 116/234

Entry S51398 from database PIR:

hypothetical protein YLR260w - yeast (*Saccharomyces cerevisiae*)

>TREMBL:SCL8479_4 gene: "YLR260W"; product: "Ylr260wp"; *Saccharomyces cerevisiae* chromosome XII cosmid 8479.

Score = 251, P = 1.0e-24, identities = 62/198, positives = 103/198

Alert BLASTP hits for DKFZphfbr2_82m6, frame 3

TREMBL:AF068749_1 gene: "SPHK1b"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1b) mRNA, complete cds., N = 2, Score = 615, P = 1.2e-92

TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1a) mRNA, partial cds., N = 2, Score = 616, P = 2e-92

TREMBL:ATF18E5_16 gene: "F18E5.160"; product: "putative protein"; *Arabidopsis thaliana* DNA chromosome 4, BAC clone F18E5 (ESSAII project), N = 2, Score = 370, P = 6.8e-33

>TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1a) mRNA, partial cds.
Length = 504

HSPs:

Score = 616 (92.4 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
Identities = 128/260 (49%), Positives = 173/260 (66%)

Query: 154 ATALTCLRLGLPLPGDGEITPDLLPRPRLLLLVNPFGRGLAWQWCKNHVLP MISEAGL 213
A C L + E LLPRP R+L+L+NP GG+G A Q ++ V P + EA +
Sbjct: 110 APVAPCQREPRDLAMEPECPGRLPRPCRVLVLLNPQGGKKGALQLFQSRVQPFLEAEI 169

Query: 214 SFNLIQTERQNHARELVQGLSLSEWDGIVTVSGDGLLHEVLNGLDRPDWEEAVKMPVGI 273
+F LI TER+NHARELV L WD + +SGDGL+HEV+NGL++RPDWE A++ P+
Sbjct: 170 TFKLILTERKNHARELVCAEELGHWDAVMSGDGLMHEVVNGLMERPDWETAIQKPLCS 229

Query: 274 LPCGSGNALAGAVNQHGGEFEPALGLDLLNCSSLLCRGGGHPLDLLSVTLASGSRCSF 333
LP GSGNALA +VN + G+E DLL+NC+LLLCR P++LLS+ ASG R +S L
Sbjct: 230 LPGSGNALAASVNHAYGYEQVTNEDLLINCTLLLCRRRLSPMNLSSLHTASGLRLYSVL 289

Query: 334 SVAWGFVSDVDIQSERFRALGSAREFTLGTVLGLATLHTYGRSLYLPA-TVEPASPTPAH 392
S++WGFV+DVD++SE++R LG RET+GT LA+L Y+G+L+YLP TV AS PA
Sbjct: 290 SLSWGFVADVDESEKYRRLGEIRFTVGTFFRLASLRIYQGQLAYLPVGTV--ASKRPAS 347

Query: 393 SL-PRAKSELTLTPDPAPPM 413
+L + + L P P +H
Sbjct: 348 TLVQKGPVDTHLVPLEEPVPSH 369

Score = 324 (48.6 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
Identities = 72/160 (45%), Positives = 100/160 (62%)

Query: 499 LPLPTPDARVGASTC---GPPDHLLPPLGTPLPPDWVTL-EGDFVLMLAISPSHLGADLV 554
LP+ T ++ AST GP D L PL P+P W + E DF+L+L + +HL ++L
Sbjct: 335 LPVGTVASKRPASTLVQKGPVDTHLVPLEEPVPSHWTVVPEQDFLLVLVLLHLSSELF 394

Query: 555 AAPHARFDDGLVHLWCWVRSGISRAALLRLFLAMERGSFSLGCPQLGYAAARAFRLEPLT 614
AAP R + G++HL +VR+G+SRALLRLFLAM++G H L CP L + AFRLEP +
Sbjct: 395 AAPMGRCEAGVMHLFYVRAGVSRALLRLFLAMQKGKHELDGPYLVHVPVVAFRLEPRS 454

Query: 615 PRGVLTVDGEQVEYGPLQAQMHPGIGTLLTGPPGCP-GRE 653
RGV +VDGE + +Q Q+HP ++ G P GR+
Sbjct: 455 QRGVFSVDGELMVCEAVQGQVHPNYLWMVCGSRDAPSGRD 494

Score = 37 (5.6 bits), Expect = 3.6e-62, Sum P(2) = 3.6e-62
Identities = 8/20 (40%), Positives = 9/20 (45%)

Query: 459 GAGDAPLSPDPLSSPPGSP 478
G+ DAP D PP P
Sbjct: 485 GSRDAPSGRDSRRGPPPEEP 504

Pedant information for DKFZphfbr2_82m6, frame 3

Report for DKFZphfbr2_82m6.3

SEQ	MNGHLEAEEQDQRPDQELTGSWGHPRSTLVRAKAMAPPPPLAASTSLHGEFGSYPA
SEGxxxxxxxxxxxxx.....
PRD	ccchhhhhhhhhccccceeeccccccccceehhhhhhhccccceeecccccccccc
SEQ	RGRPFALTITSQALHIQRLRPKPEARPRGGLVPLAEVSGCCTLRSRSPSDSAAFYCIYTY
SEG
PRD	ccccceehhhhhhhhhhhhhhhccccccccccccceeeeeceeeeeccccceeeeeee
SEQ	PRGRRGARRRATRTFRADGAATYEENRAEAQRWATALTCLLRGLPLPGDGEITPDLLRP
SEG	.xxxxxxxxxxxxxxxxxxxxx.....xxxxx
PRD	ccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc
SEQ	PRLLLLVNPFGRGLAWQCKNHVLPIMISEAGLSFNLIQTERQNHARELVQGLSLSEWDG
SEG	xxxxxx
PRD	ceeeeeccccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhccccce
SEQ	IVTVSGDGLLHEVLNGLDRPDWEEAVKMPVGILPCGSGNALAGAVNQHGGFEPALGLDL
SEGxxxxx
PRD	eeeeccccceeeccccccccchhhhhccccceeeccccccccccccccccccccccccchhhhhh
SEQ	LLNCSLLRCGGGHPDLLSVTLASGSRCSFSLVSWGFEVSDVDIQUERFRALGSARFTL
SEG	xxxxxxxxxxxxxxxx
PRD	hhhhhhccccccccccccceeeeeccccceeeeeeeccccceeehhhhhhhhhhhhhhhhc
SEQ	GTVLGLATLHTYRGRLSYLPATVEPASPTPAHSLPRAKSELTLTPDPAPPMHSLPHRSV
SEG
PRD	hhhhhhhhhhhhcc
SEQ	SDLPLPLPQPALASPGSPEPLPILSLNGGPELAGDWGGAGDAPLSPDLLSSPPGSPKA
SEG	.xxxxxxxxxxxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD	ccccccccccccccccccccceeeccccccccccccccccccccccccccccccccccccce
SEQ	ALHSPVSEGAPIVPSSGLPLTPDARVGASTCGPPDHLLPLPLGTPLPPDWVTLEGDFVL
SEG	xx.....xxxxxxxxxxxxxx
PRD	eeccce
SEQ	MLAISPSHLGADLVAAPHARFDDGLVHLCWVRSGISRAALLRLFLAMERGSFSLGCPQL
SEG
PRD	eeeeccccccccccccccccccccceeeeeeeccchhhhhhhhhhhhhhhccccceeeccch
SEQ	GAAARAFRLEPLTPRGVLTVDGEQVEYGLPQAQMHPIGITLLTGPPGCPGREP
SEGxxxxxxxxxxxxxxxxxxxx
PRD	hhhhhhhhhhccccccccceeeccccceccccccccccccccccceeecccccccccccc

Prosites for DKFZphfbr2_82m6.3

PS00001	303->307	ASN_GLYCOSYLATION	PDOC00001
PS00002	245->249	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	129->133	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	102->105	PKC_PHOSPHO_SITE	PDOC00005
PS00005	134->137	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	347->350	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	371->374	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	614->617	PKC_PHOSPHO_SITE	PDOC00005
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006

PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	341->345	CK2_PHOSPHO_SITE	PDOC00006
PS00006	419->423	CK2_PHOSPHO_SITE	PDOC00006
PS00007	106->115	TYR_PHOSPHO_SITE	PDOC00007
PS00008	56->62	MYRISTYL	PDOC00008
PS00008	212->218	MYRISTYL	PDOC00008
PS00008	232->238	MYRISTYL	PDOC00008
PS00008	272->278	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	279->285	MYRISTYL	PDOC00008
PS00008	361->367	MYRISTYL	PDOC00008
PS00008	476->482	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	574->580	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	640->646	MYRISTYL	PDOC00008
PS00009	122->126	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_82m6.3)

DKFZphfkd2_lj9

group: kidney derived

DKFZphfkd2_lj9.3 encodes a novel 105 amino acid protein with high similarity to *Xenopus laevis* XLCL2 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

strong similarity to XLCL2 protein, African clawed frog

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 2955 bp

Poly A stretch at pos. 2935, polyadenylation signal at pos. 2915

```
1 GGGGGGGGCT GAGTGCTCAG TGGAGAGCGG GGAGTTGTGT CCACCTTGCC
51 GACGTGCGCTA GCCGTGGGGC TGTCCTGGGA AGGCGGACGG CGAGCGCCCCG
101 GTGTCCGCAC TCGGCGCCCT GCCGTGCCCG TCTGCGCCCG TGTCATCCTC
151 ACTCGGGACG CAGGGACCGT TTTTAAATCA CAGGGGCGTG TGTCAGCCTG
201 CCCTAGGACT TCATGTCTAT ATATTTCCTC ATTCAGTCCG CCGACTATCT
251 GAGATCGGGC AAGATGACTG AGGTGATGAT GAACACCCAG CCCATGGAGG
301 AGATCGCCCT CAGCCCCCGC AAGGATGGCC TTCTCTACCA GATCTTCCCA
351 GACCCGCTCAG ATTTTGACCG CCGCTGCAAA CTGAAGGACC GTCTGCCCTC
401 CATAGTGGTG GAACCCACAG AAGGGGAGGT GGAGAGCGGG GAGCTCCGGT
451 GGCCCCCTGA GGAGTTCCCTG GTCCAGGAGG ATGAGCAAGA TAACTGCGAA
501 GAGACAGCGA AAGAAAATAA AGAGCAGTAG AGTCCCTGTG GACTCCCATG
551 GGTTCATACCA GCCAGCATCT GTTCTGAAAC TGTGTTTTTC CCATCATGAC
601 GGAAGAAGAG AGTGAGCCGC AATTGTTCTG AAAATGTCAA ACGAGGCTTC
651 TGTTTTGCAC CTGCAGATCA CCGAGTTGGT TTTCTTTTCT TTTCTTGCC
701 TTTTTTTTTT TTTGAAATTT GCCGAGCAGT GGAGCCCTCT GACAATTTGC
751 AAGGCCCTCT GAGAAAGGAA GCTGCTTAGA GCCAGGGGGT TAGTGGGTGA
801 GGGGAGCGAG TGCTGTTTTT GAGATCATTA TCTGAACTCA GGCAGCCTAG
851 TAGAGGCAGT GGTGGGATTC CAATGGGTCT TGGTGGGTGG GAGGTGGGGC
901 ATGTGCAAG CAAGCAAGGA ACATTGGGGG TAAGAAAACA AACATGAGGC
951 AAAAGAAAAA ATACATGTTT TTAAGAAAAC ATTGAGCAGA GAACTGCAGC
1001 CAGGATGCCG TCAGCAGACA TTCACCTCTG CCGCTGGGAC ATCAGAAAAC
1051 AAAGTCTTCA TCTCTCTCTC CAGTTTCACC CACCCACCCC TTTGCTTTCA
1101 TTTCAAGGTG GTTGGTCTAT ATGACAGGGA GGAGAGTAAA GGAGAGCAGG
1151 AGCAATTGGC TGCCCTGCAA GCCAGCTGGA GGTGAAGTGC AGGAAAGGAA
1201 AGGTCAACCC ATTCTACTCC ATGGCCTCTC TGCTCCCAGC TGTGGTAGGC
1251 TCACATAGCC AGTGTGATCG GTTTTTAAGA GGCAGTGCTT TTCAGCTTTT
1301 CTCCTGATA TATCCATTTT GCTTCCCAGC ACTTTTATAG AGTAGTGAGA
1351 GCATCTCCTG CCCTTGTTGG AAGCCCCAGG GTGGACACTC AGCACGAAGG
1401 TCTCTCCCTT AACTGCTGCC CTTCCAAGAC TTGCTCCCGA GATGGAGTGG
1451 GCGTGGTCTT CCAGGCTGGC CTTCTCTCTT CCTCAGCCGC ACCTTCCCTG
1501 CCCCAGCCCT AGCAGCCATG GGTACATGGG TCCCAGCTC ACCTATGGAT
1551 TCCCAGCAGT CTGCCAGCT GCAGTACTCA CGCCCATG GGGATCTTGG
1601 TCTGTTTTTC TTGTGGGAGC CTAGTGGAGA GCAGAGCTGG CTTTTTATGT
1651 GTCTTGTTGG GGAGGTGACT TGCATGGTGG GGACAAGGCT GTCGTGGCAA
1701 CCTTGGGATC GAGTTTGAGA CTAAGGATG TCATGAGATC CCTGGCTTCT
1751 CCCCATGTTG TTCCCGGACA AGGGCAGAAG GGAGGCATGG CAAGGGACCT
1801 CTGCTGTCTT TACTCAACAG TGGTCCTCAT CCCTCCCCAC CTCCCAGTGC
1851 TTCCTGCAAG GGCACCAAGT GTATGAGAAA GTTGGCTTTT GGACTTAGGA
1901 TTTCTTATTG TAGCTAAGAG CCATCTGAAG CAGCAGGTTG CAGGACAAAT
1951 GCTTCAGTCC GCCGAGAGCA GTACCGTGTG GCCAAGAGGT GGAAGCAGAG
2001 CCTTCTTGA GCTAAACTCG GCCAACAAG GCACGCAGCA TGTCCCCTCA
2051 GGTCTCCAGT CAGTCCAGGT TGACCCTCAG TTCTGGACGT GTGTATATAG
2101 CTGTATTTAA TACCTCAAGG TCATTGTGGC TCTGGGGATG CCAGGGCAGG
2151 AGGAGCAGGG TGCGCTGTGG ACACAGCAGT CCGCGGAATT CCGTTCTGGG
2201 AAGCCAAATG TCGCCGGCAC CCCTTGCTTC CTCCCTCTGT TGTCTGCCTG
2251 TGTGACACAC ATCAATGGCA ATAACCTCTT CCAACTCTCT GCAGAAGTGG
2301 GAGAGGCGCG CAGCCTGCAC CGAGAGGGGC TTCTCTCTCT CTGTGCTCCC
2351 GCTTCGTTCT GTTTTGGCTG CAGAGAGTGG TTCATCCATA CTCTCATTCC
2401 CTCGCCCTCC CTTGTGGACG GGGGTCTTGC CTTTTCAATT CCTGTGTTTT
2451 GGTGTCTTCC CTTATCTGCT ACCCTGAAAT ACCTGTCCCT GTCTTGTCTG
2501 GTGATGGGAA CATGCTTGTA AACTGCGTAA CAAATCTACT TTGTGTATGT
2551 GTCTGTTTAT GGGGGTGGTT TATTATTTTT GCTGGTCCCT AGACCACTTT
2601 GTATGACCTT TTGCAGTCTG AGCAGGCCAG GGGCTGACAG CTAATGTCAG
2651 GACCCTCAGC GGTGGAGCCT GCTGGGGGGA CCCAGCTGCT CTTGGACAAG
```



```
2701 TGGCTGAGCT CCTATCTGGC CTCCTCTTTT TTTTTTTTTT CAAGTAATTT
2751 GTGTGTATTT CTAACGTATT GTATTGAAAA AATTCCTAGT ATTTTCAGTAA
2801 AAATGCCGTGT TGTGAGATGA ACCTCCTGTA ACTTCTATCT GTTCTTTTTT
2851 GAGGCTCAGG GAGAACTAG CATTTTTTTT TTTCCAAACT ACTTTTTGTC
2901 ACTGTGACAG TTGTAAATAA AGTTTGAAAA TGCTCAAAAA AAAAAAAAAA
2951 AAAAC
```

BLAST Results

Entry HSG19750 from database EMBL:
human STS A001X24.
Score = 1050, P = 1.9e-39, identities = 212/213

Entry HSG20267 from database EMBL:
human STS A005C12.
Score = 610, P = 4.1e-19, identities = 122/122

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 213 bp to 527 bp; peptide length: 105
Category: strong similarity to known protein
Classification: unset

```
1 MSYFPIHCP DYLRSAKMTE VMNTQPMEE IGLSPRKDGL SYQIFPDPSD
51 FDRRCKLKDR LPSIVVEPTE GEVESGELRW PPEEFLVQED EQDNCEETAK
101 ENKEQ
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_lj9, frame 3

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8e-42

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8.2e-42

>PIR:S52241 XLCL2 protein - African clawed frog
Length = 102

HSPs:

Score = 443 (66.5 bits), Expect = 8.0e-42, P = 8.0e-42
Identities = 80/104 (76%), Positives = 95/104 (91%)

```
Query: 1 MSYFPIHCPDYLRSAKMTEVMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRRCKLKDR 60
      MS+++PIHC DYLRSA+MTEV+MNTQ M+EIGLSPRKD SYQIFPDPSDF+R CKLKDR
Sbjct: 1 MSVFYPIHCTDYLRSAEMTEVIMNTQSMDEIGLSPRKD--SYQIFPDPSDFERCCKLKDR 58

Query: 61 LPSIVVEPTEGEVESGELRWPPPEEFLVQEDEQDNCEETAKENKE 104
      LPSIVVEPTEG+VESGELRWPPPEEF+V ED++ C++T KEN++
Sbjct: 59 LPSIVVEPTEGDVESGELRWPPPEEFVVDDEKGTCDQTKKENEQ 102
```

Pedant information for DKFZphfkd2_lj9, frame 3

Report for DKFZphfkd2_lj9.3

```
[LENGTH] 105
[MW]      12269.78
[PI]      4.40
[PMOL]    PIR:S52241 XLCL2 protein - African clawed frog 5e-44
```

[KW] Alpha_Beta

SEQ MSIYFPIHCPDYLRSAKMTVMMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRRCKLKDR
PRD cccccccccchhhhhhhhhhhccccccccccccccccccccccccccccccccchhhhhhhc

SEQ LPSIVVEPTEGEVESGELRWPPEEFLVQEDEQDNCEETAKENKEQ
PRD cccccccccccccccccccccccccccccccccccccchhhhhhhhhccc

(No Prosite data available for DKFZphfd2_1j9.3)

(No Pfam data available for DKFZphfd2_1j9.3)

DKFZphfkd2_24a15

group: transmembrane protein

DKFZphfkd2_24a15 encodes a novel amino acid protein with similarity to C. elegans cosmid R07G3.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to C. elegans R07G3.8

membrane regions: 1

Summary DKFZphfkd2_24a15 encodes a novel 323 amino acid protein, with similarity to C. elegans R07G3.8.

similarity to C. elegans R07G3.8

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1513 bp

Poly A stretch at pos. 1494, no polyadenylation signal found

```
1  GGGGTA CTG GCGGCGGCGG AGCGGGCGGC AGAGCAGGGC GCGGCGGACT
51  CGCAGGGTAC CACCATCTTA AGGACAGAAA AGCTACAGGA CTCTAGGAGG
101 CCACCGTCTT GATTTGGGAA GTCCAACTTA CTTTGCCAG ACAGCAGCTA
151 AGCTGGTTCA TCCCATCAGC CTGGATTGGT GAAACTGAAT CACAGGAGAT
201 ATTTCCAGGT TTGCTGGGAT GGGAAACCTG CTCAAAGTCC TTACCAGGGA
251 AATTGAAAC TATCCACACT TTTCTCTGGA TTTTGAAAT GCTCAGCCTA
301 CAGAAGGAGA GAGAGAAATC TGGAAACCAGA TCAGCGCCGT CCTTCAGGAT
351 TCTGAGAGCA TCCTTGCAGA CCTGCAGGCT TACAAAGGCG CAGGCCGAGA
401 GATCCGAGAT GCAATTCAAA ATCCCAATGA CATTGAGCTT CAAGAAAAAG
451 CTTGGAATGC GGTGTGCCCT CTGTTGTGA GGCTAAAGAG ATTTTACGAG
501 TTTTCCATTA GACTAGAAAA AGCTCTTCAG AGTTTATTGG AATCTCTGAC
551 TTGTCCACCC TACACACCAA CCAACACCTT GGAAAGGGAA CAGGCCCTGG
601 CAAAGGAGTT TGCCGAAATT TTACATTTTA CCCTTCGATT CGATGAGCTG
651 AAGATGAGGA ACCCGGCTAT TCAGAATGAC TTCAGCTACT ACAGAAGAAC
701 AATCAGTCCG AACCGCATCA ACAACATGCA CCTAGACATT GAGAATGAAG
751 TCAATAATGA GATGGCCAAT CGAATGTCCC TCTTCTATGC AGAAGCCACG
801 CCAATGCTGA AAACCTTAG CAATGCCACA ATGCACTTTG TCTCTGAAAA
851 CAAAACCTCG CCAATAGAGA ACACCACAGA CTGCCTCAGC ACAATGACAA
901 GTGTCTGTAA AGTCATGCTG GAAACTCCGG AGTACAGAAG TAGGTTTACG
951 AGTGAAGAGA CCCTGATGTT CTGCATGAGG GTGATGGTGG GAGTCATCAT
1001 CCTCTATGAC CATGTCCACC CTGTGGGAGC TTTCTGCAAG ACATCCAAGA
1051 TCATATGATA AGGCTGCATA AAAGTTTGA AGGAGCAGGC CCCAGACAGT
1101 GTGGAGGGGC TGCTAAATGC CTTGAGGTTT ACTACAAAGC ACTTGAAACGA
1151 TGAATCAACT TCCAAACAGA TTCGAGCAAT GCTTCAGTAG AGCTCTGCTC
1201 AAAGAAGAGG ATCTATGTGC TGACCTCAGA AGATGTATAT GTTTACATAA
1251 TTTAATACAG ATTGATGTGA ATACTTGTGT ATTTACATAA CCGTTTCCTT
1301 CTTGTCACTG AAATATATGG ACCTTAATT GTATCCTGAC TGAATCAACC
1351 CAGCAGAGCA TAAATTGACT TGAGAGCCTT ACCTTTGATG TCTGAAATGA
1401 AACCCCTTTC TCCAAAGGCA AAATTCGGAG ACTTTGATCT TTGCTACTGG
1451 AGTCCCTTAA CAACATCTAT AACGATAAAA AATTCCTAAT TGTCAAAAAA
1501 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

 ORF from 219 bp to 1187 bp; peptide length: 323
 Category: similarity to unknown protein

```

1 MGNNLLKVLTR EIENYPHFLL DFENAPQTEG EREIWNQISA VLQDSESILA
51 DLQAYKGAGP EIRDAIONPN DIQLQEKAWN AVCPLVVRLK RFYEFSIRLE
101 KALQSLLES LCPPTPTQH LEREQALAKE FAEILHFTLR FDELKMRNPA
151 IQNDFSYYRR TISRNRINNM HLDIENEVNN EMANRMSLFY AEATPMLKTL
201 SNATMHFVSE NKTLPIENTT DCLSTMTSVC KVMLETPEYR SRFTSEETLM
251 FCMRVMVGVI ILYDHVHPVG AFCKTSKIDM KGCIKVLKEQ APDSVEGLLN
301 ALRFTTKHLN DESTSKQIRA MLQ
  
```

BLASTP hits

Entry CER07G3_7 from database TREMBL:
 gene: "R07G3.8"; Caenorhabditis elegans cosmid R07G3.
 Score = 544, P = 1.4e-52, identities = 119/323, positives = 186/323

Alert BLASTP hits for DKFZphfd2_24a15, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfd2_24a15, frame 3

Report for DKFZphfd2_24a15.3

[LENGTH] 323
 [MW] 37313.06
 [pI] 5.71
 [HOMOL] TREMBL:CER07G3_7 gene: "R07G3.8"; Caenorhabditis elegans cosmid R07G3. 4e-54

[PROSITE] MYRISTYL 1
 [PROSITE] CK2_PHOSPHO_SITE 4
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 5
 [PROSITE] ASN_GLYCOSYLATION 3
 [KW] TRANSMEMBRANE 1

```

SEQ  MGNNLLKVLTR E IENYPHFLL DFENAPQTEG EREIWNQISA VLQDSESILA DLQAYKGAGP
PRD  cccccchhhhhhhccccccccccccchhhhhhhhhhhhhhhcchhhhhhhhhcccccc
MEM  .....

SEQ  EIRDAIONPNDIQLQEKAWN AVCPVVRLK RFYEFSIRLE KALQSLLES LCPPTPTQH
PRD  hhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhh
MEM  .....

SEQ  LEREQALAKE FAEILHFTLR FDELKMRNPA IQNDFSYYRR TISRNRINNM HLDIENEVNN
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhccchhhhhhhhhhhhhhh
MEM  .....

SEQ  EMANRMSLFY AEATPMLKTL SNATMHFVSE NKTLPIENTT DCLSTMTSVC KVMLETPEYR
PRD  hhhhhhhhhhhhhccccchhhhhhhcccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  SRFTSEETLM FCMRVMVGVI ILYDHVHPVG AFCKTSKIDM KGCIKVLKEQ APDSVEGLLN
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ  ALRFTTKHLN DESTSKQIRA MLQ
PRD  hhhhhhccccccccchhhhhhhccc
MEM  .....
  
```

Prosites for DKFZphfd2_24a15.3

PS00001	202->206	ASN_GLYCOSYLATION	PDOC00001
PS00001	211->215	ASN_GLYCOSYLATION	PDOC00001
PS00001	218->222	ASN_GLYCOSYLATION	PDOC00001
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	275->278	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005

PS00005	314->317	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00007	231->240	TYR_PHOSPHO_SITE	PDOC00007
PS00008	297->303	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfd2_24a15.3)

DKFZphfkd2_24b15

group: metabolism

DKFZphfkd2_24b15 encodes a novel 612 amino acid protein with similarity to bacterial and yeast phosphoglucomutase and phosphomannomutases.

The novel protein contains a phosphoserine signature typical for phosphoglucomutase (EC 5.4.2.2) or phosphomannomutase (EC 5.4.2.8). Thus, the protein seems to be taking part in the conversion of hexose phosphates.

The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

similarity to phosphomannomutases

complete cDNA, complete cds, EST hits
potential start at bp 30 matches kozak consensus PyCnatgG,

Sequenced by GBF

Locus: map="158.8 cR from top of Chr4 linkage group"

Insert length: 2204 bp

Poly A stretch at pos. 2186, no polyadenylation signal found

```
1 GGGCTCTGCA GCGGTAGCAC AAGCTCAGCG ATGGCGGCTC CAGAAGGCAG
51 CGGTCTAGGC GAGGACGCCC GGCTGGACCA GGAGACCGCC CAGTGGCTGC
101 GCTGGGACAA GAATTCCTTA ACTTTGGAGG CAGTGAAACG ACTAATAGCA
151 GAAGGTAATA AAGAAGAACT ACGAAAATGT TTTGGGGCCC GAATGGAGTT
201 TGGGACAGCT GGCCTCCGAG CTGCTATGGG ACCTGGAATT TCTCGTATGA
251 ATGACTTGAC CATCATCCAG ACTACACAGG GATTTTGAG ATACCTGGAA
301 AAACAATTCA GTGACTTAAA GCAGAAAGGC ATCGTGATCA GTTTTGACGC
351 CCGAGCTCAT CCATCCAGTG GGGGTAGCAG CAGAAGGTTT GCCCGACTTG
401 CTGCAACCCAC ATTTATCAGT CAGGGGATTC CTGTGTACCT CTTTCTGAT
451 ATAACGCCAA CCCCTTTTGT GCCCTTCACA GTATCACATT TGAAACTTTG
501 TGCTGGAATC ATGATAACTG CATCTCACA TCCAAAGCAG GATAATGGTT
551 ATAAGGTCTA TTGGGATAAT GGAGCTCAGA TCATTCTCTC TCACGATAAA
601 GGGATTTCTC AAGCTATTGA AGAAAATCTA GAACCGTGGC CTCAGCTTG
651 GGACGATTCT TTAATTGATA GCAGTCCACT TCTCCACAA CCGAGTGCTT
701 CCATCAATAA TGACTACTTT GAAGACCTTA AAAAGTACTG TTTCCACAGG
751 AGCGTGAACA GGGAGACAAA GGTGAAGTTT GTGCACACCT CTGTCCATGG
801 GGTGGGTCTAT AGCTTTGTGC AGTCAGCTTT CAAGGCTTTT GACCTTGTTT
851 CTCCTGAGGC TGTTCCTGAA CAGAGAGATC CGGATCCTGA GTTTCCAAAC
901 GTGAATACC CGAATCCCGA AGAGGGGAAA GGTGTCTTGA CTTTGTCTTT
951 TGCTTTGGCT GACAAAACCA AGGCCAGAA TGTTTTAGCT AACGACCCGG
1001 ATGCTGATAG ACTTGCTGTG GCAGAAAAGC AAGACAGTGG TGAATGGAGG
1051 GTGTTTTTCA GCAATGAGTT GGGGGCCCTC CTGGGCTGGT GGCTTTTTTAC
1101 ATCTTGGAAA GAGAAGAACC AGGATCGCAG TGCTCTCAA GACACGTACA
1151 TGTTGTCCAG CACCGTCTCC TCCAAAATCT TGCGGGCCAT TGCCTTAAAG
1201 GAAGGTTTTT ATTTTGAGGA AACATTAAC TGCCTTAAGT GGATGGGAAA
1251 CAGAGCCAAA CAGCTAATAG ACCAGGGGAA AACTGTTTTA TTTGCATTTG
1301 AAGAGCTAT TGGATACATG TGCTGCCCTT TTGTTCTGGA CAAAGATGGA
1351 GTCAGTGCCG CTGTCATAAG TGCAGAGTTG GCTAGCTTCC TAGCAACCAA
1401 GAATTTGTCT TTGTCTCAGC AACTAAAGGC CATTATATGT GAGTATGGCT
1451 ACCATATTAC TAAAGCTTCC TATTTTATCT GCCATGATCA AGAAACCAT
1501 AAGAAATTAT TTGAAAACCT CAGAAACTAC GATGGAAAAA ATAATTATCC
1551 AAAAGCTTGT GGCAAATTG AAATTTCTGC CATTAGGGAC CTTACAACATG
1601 GCTATGATGA TAGCCAACCT GATAAAAAAG CTGTTCTTCC CACTAGTAAA
1651 AGCAGCCAAA TGATCACCTT CACCTTTGCT AATGGAGGCG TGGCCACCAT
1701 GCGCACCAGT GGGACAGAGC CCAAAATCAA GTACTATGCA GAGCTGTGTG
1751 CCCCACTTGG GAACAGTGAT CCTGAGCAGC TGAAGAAGGA ACTGAATGAA
1801 CTGGTCAGTG CTATTGAAGA ACATTTTTC CAGCCACAGA AGTACAATCT
1851 GCAGCCAAAA GCAGACTAAA ATAGTCCAGC CTGGGTATA CTTGCATTTA
1901 CTTACAATTA AGCTGGGTTT AACTTGTTAA GCAATATTTT TAAGGGCCAA
1951 ATGATTCAAA ACATCACAGG TATTTATGTG TTTTACAAAG ACCTACATTC
2001 CTCATTGTTT CATGTTTGAC CTTTAAAGTG AAAAAAGAAA ATGGCCAAAC
2051 CCAACAAACT AACATTCCCTA CTAAAAAGTT GAGCTTGAC ATATTTTGAA
2101 TTTTGTAAAG TGAAGATTTT TAAACTGACT AACTTAAAAA AATAGATTGT
2151 AATTGATGTG CCTTAATTG CATAAATCAT AAATGTAAAA AAAAAAAA
2201 AAAA
```

BLAST Results

Entry HS705145 from database EMBL:

human STS WI-6820.

Score = 1261, P = 3.6e-52, identities = 253/254

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 31 bp to 1866 bp; peptide length: 612
Category: strong similarity to known protein

```

1  MAAPEGSGGLG EDARLDQETA QWLRLWDKNSL TLEAVKRLIA EGNKEELRKC
51 FGARMEFGTA GLRAAMGPGI SRMNDLTIIQ TTQGFRCYLE KQFSDLKQKG
101 IVISFDARAH PSSGGSSRRF ARLAATTFIS QGIPVYLFSD ITPTPFVPFT
151 VSHLKLKAGI MITASHNPKQ DNGYKVYWDN GAQIISPHDK GISQAIENL
201 EPWPQAWDDS LIDSSPLLHN PSASINNDYF EDLKKYCFHR SVNRETKVKF
251 VHTSVHGVGH SFVQSFAKAF DLVPPEAVPE QRPDPPEFPT VKYPNPEEGK
301 GVLTLSFALA DTKKARIVLA NDPDADRLAV AEKQDSGEWR VFSGNELGAL
351 LGWNLFTSWK EKNQDRSALK DTYMLSSTVS SKILRAIALK EGFHFEETLT
401 GFKWMGNRAK QLIDQGKTIVL FAFEEAIGYM CCPFVLDDKG VSAAVISAEI
451 ASFLATKNLS LSQQLKAIYV EYGVHITKAS YFICHQDQETI KKLLENLRNY
501 DGKNNYPKAC GKFEISAIRD LTTGYDDSQP DKKAVLPTSK SSQMIFTFFA
551 NGGVATMRTS GTEPKIKYYA ELCAPPGNSD PEQLKKELNE LVSAIEEHFF
601 QPQKYNLQPK AD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24b15, frame 1

TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B, N = 1, Score = 1431, P = 1.6e-146

TREMBL:SPCC1840_5 gene: "SPCC1840.05c"; product: "similarity to phosphomannomutases"; S.pombe chromosome III cosmid c1840., N = 1, Score = 1210, P = 4.2e-123

PIR:S54585 hypothetical protein YMR278w - yeast (Saccharomyces cerevisiae), N = 1, Score = 1046, P = 1e-105

PIR:A71299 probable phosphomannomutase (manB) - syphilis spirochete, N = 1, Score = 697, P = 9.7e-69

>TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B
Length = 595

HSPs:

Score = 1431 (214.7 bits), Expect = 1.6e-146, P = 1.6e-146
Identities = 285/598 (47%), Positives = 393/598 (65%)

```

Query:   13  ARLDQETAQWLRLWDKNSLTLEAVKRLIAEGNKEELRKC FGARMEFGTAGLRAAMGPGISR 72
          A+LD++ A WL WDKN      +++L+ E N + L+      R+ FGTA+R+ M G R
Sbjct:   6  AKLDKQVADWLAWDKNDKRNNEIQKLVDENVDALKARMDTRLVFGTAGVRSMPMAGFGR 65

Query:   73  MNDLTIIQTTQGFRCYLEKQFSDLKQKGVISFDARAH PSSGGSSRRFARLAATTFISQG 132
          +NDLTIIQ T GF R++   + K G+ I FD R +      SRRFA L+A F+
Sbjct:   66  LNDLTIIQITHGFARHMLNVYGQPKN-GVAIGFDGRYN-----SRRFAELSANVFVRNN 118

Query:   133 IPVYLFSDITPTPFVPFTVSHLKLKAGIMITASHNPKQDNGYKVYWDN GAQIISPHDKGI 192
          IPVYLF+++PTP V +   L AG++ITASHNPK+DNGYK YW NGAQII PHD I
Sbjct:   119 IPVYLFSEVSPTPVVSWATIKLGCDAGLIITASHNPKEDNGYKAYWSNGAQIIGPHDTEI 178

Query:   193 SQAIENLEPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHRSVNRETKVKFVH 252
          + E +P + WD S + SSPL H+   I+ YFE K F R +N T +KF +
Sbjct:   179 VRIKEAEPQPRDEYWDLSSELKSSPLFHSADVVVID-PYFEVEKSLNFTREINGSTPLKFTY 237

Query:   253 TSVHGVGHGSFVQSFAKAFDLVPPE--AVPEQRDPDPPEFPTVKYPNPEEGKGVLTLSFALA 310
          ++ HG+G+ + + F F      +V EQ+DP+P+FPT+ +PNPEEG+ VLT+ A
Sbjct:   238 SAFHGIGYHYTKRMFAEFGFPASSFISVAEQQDPNPDFTPTIPFPNPEEGRKVLTLMETA 297

```

Query: 311 DTKARIVLANDPDADRLAVAQKQDSGEWRVFSGNELGALLGWWLFTSWKEKNQDRSALK 370
 DK + ++LANDPDADR+ +AEKQ GEWRVF+GNE+GAL+ WW++T+W++ N + A K
 Sbjct: 298 DKNGSTVILANDPDADRIQMAEKQKDGGEWRVFTGNEMGALITWWIWTNRKANPNADASK 357

Query: 371 DTYMLSSSTVSSKILRAIALKEGFHFEETLTGFKWMGNRAKQLIDQKTVLFAFEEAIGYM 430
 Y+L+S VSS+I++ IA EGF E TLGFKWMGNRA++L G V+ A+EE+IGYM
 Sbjct: 358 -VYILNSAVSSQIVKTIADAEGFKNETTLTGFKWMGNRAEELRADGNQVILAWEEISIGYM 416

Query: 431 CCP-FVLDDKGVSAAVISAELASFLATKNLSLSQLKAIYVEYGYHITKASYFICHQDET 489
 P +DKDGVSA + AE+A+FL + SL QL A+Y YG+H+ +++Y++ E
 Sbjct: 417 --PGHTMDKDGVSAAVFAEIAAFLHAEGKSLQDQLYALYNRYGFHLVRSTYWMVPAPEV 474

Query: 490 IKKLFENLRNYDGKNYPKACGKFEISAIRDLTTGYDDSQPDKKAVLPTSXSQMITFTF 549
 KKLF LR D K +P G+ E++++RDLT GYD+S+PD K VLP S SS+M+TF
 Sbjct: 475 TKKLFSTLRA-DLK--FPTKIGAEVASVRDLTIGYDNSKPDNKPVLPLSTSEMVTFFL 531

Query: 550 ANGGVATMRTSGTEPKIKYYAELCAPPGNS--DPEQLKKELNELVSAIEEHFFQPKYNL 607
 G V T+R SGTEPKIKYY EL PG + D E + E++L + +PQ++ L
 Sbjct: 532 KTGSVTTLRASGTEPKIKYYIELITAPGKTQNDLESVISEMDQLEKDVVATLLRPQQFGL 591

Query: 608 QPK 610
 P+
 Sbjct: 592 IPR 594

Pedant information for DKFZphfd2_24b15, frame 1

Report for DKFZphfd2_24b15.1

[LENGTH] 612
 [MW] 68311.58
 [pI] 6.28
 [HOMOL] TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B 1e-157

[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YMR278w] 1e-111
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI0740] 3e-66
 [FUNCAT] c energy conversion [M. genitalium, MG053] 4e-50
 [FUNCAT] m outer membrane and cell wall [H. influenzae, HI1463] 2e-04
 [BLOCKS] BL00607D cAMP phosphodiesterases class-II proteins
 [BLOCKS] BL00710 Phosphoglucosyltransferase and phosphomannomutase phosphoserine signa
 [EC] 5.4.2.8 Phosphomannomutase 3e-56
 [EC] 5.4.2.2 Phosphoglucosyltransferase 1e-09
 [PIRKE] isomerase 3e-56
 [PIRKE] intramolecular transferase 3e-56
 [SUPFAM] Methanobacterium thermoautotrophicum phosphomannomutase 1e-06
 [SUPFAM] probable phosphorylating protein ureC 9e-06
 [PROSITE] PGM_PMM1
 [PROSITE] MYRISTYL 10
 [PROSITE] LIPOCALIN 2
 [PROSITE] CK2_PHOSPHO_SITE 9
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Phosphoglucosyltransferase and phosphomannomutase phosphoserine
 [KW] Alpha_Beta

SEQ MAAPEGSGLGEDARLDQETAQWLRWDKNSLTLEAVKRLIAEGNKEELRKCFGARMEFGTA
 PRD ccc

SEQ GLRAAMGPGISRMNDLTIIQTTQGFRCYLEKQFSDLKQKIVISFDARAHPSGGSSRRF
 PRD ccc

SEQ ARLAATTFISQIGIPVYLFSDITPTFPVFTVSHLKLACAGIMITASHNPKQDNGYKVVWDN
 PRD hhhhhhhhhhhcc

SEQ GAQIISPHDKGISQAIENLEFPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHR
 PRD ccc

SEQ SVNRETKVKFVHTSVHGVGHSFVQSAFAFDLVPPEAVPEQRDPDPEFPTVKYPNPEEGK
 PRD ccc

SEQ GVLTLSPALADTKARIVLANDPDADRLAVAQKQDSGEWRVFSGNELGALLGWWLFTSWK
 PRD hhhhhhhhhhhhhcc

SEQ EKNQDRSALKDTYMLSSSTVSSKILRAIALKEGFHFEETLTGFKWMGNRAKQLIDQKTVL
 PRD hcc


```

SEQ  FAFEEAIGYMCCPFVLDKDGVSAAVISAEASFLATKNLSLSQQLKAIYVEYGYHITKAS
PRD  hhhhhccccccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhhhcccccccccc

SEQ  YFICHQDETIKKLFENLRNYDGKNYPKACGKFEISAIRDLTTGYDDSQPDKKAVLPTSK
PRD  eeeccchhhhhhhhhhhhhccccccccchhhhhhhcccccccccccccccccccccc

SEQ  SSQMITFTFANGGVATMRSTSGTEPKIKYYAELCAPPGNSDPEQLKKELNELVSAIEEHFF
PRD  cccceeeecccccceeeecccccccccceeeecccccchhhhhhhhhhhhhhhhhhhhh

SEQ  QPQKYNLQPKAD
PRD  cccccccccccc

```

Prosites for DKFZphfd2_24b15.1

PS00001	458->462	ASN_GLYCOSYLATION	PDOC00001
PS00002	7->11	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	117->120	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	380->383	PKC_PHOSPHO_SITE	PDOC00005
PS00005	489->492	PKC_PHOSPHO_SITE	PDOC00005
PS00005	538->541	PKC_PHOSPHO_SITE	PDOC00005
PS00005	556->559	PKC_PHOSPHO_SITE	PDOC00005
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	343->347	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	528->532	CK2_PHOSPHO_SITE	PDOC00006
PS00006	560->564	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	61->67	MYRISTYL	PDOC00008
PS00008	100->106	MYRISTYL	PDOC00008
PS00008	159->165	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	257->263	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00008	348->354	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00710	159->174	PGM_PMM	PDOC00589
PS00213	346->358	LIPOCALIN	PDOC00187
PS00213	344->358	LIPOCALIN	PDOC00187

Pfam for DKFZphfd2_24b15.1

HMM_NAME	Phosphoglucosyltransferase and phosphomannosyltransferase phosphoserine		
HMM	*GvnVidIGQNGMMPTPMIYFaIRTYKhmcggGIMITaSHNPGGPDnDN		
	G+ V +	++PTP + F +	H+++ +GIMITASHNP DN
Query	132	GIPVYLFS--DITPTPFVPFTVS---HLKLCAGIMITASHNP--KQ-DN	172
HMM	GIK*		
	G+K		
Query	173	GYK	175

DKFZphfkd2_24e23

group: kidney derived

DKFZphfkd2_24e23 encodes a novel 198 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit,
many ATGs in front of the ORF

Sequenced by GBF

Locus: unknown

Insert length: 1723 bp

Poly A stretch at pos. 1695, no polyadenylation signal found

```
1 GGGGGATTTT CGATCATGAC AACGATAGCA ATTGATATAC CTTCAAAATA
51 CGTGTCCAGT GAGTGTGAT TGTGTGTGGT TTCTCTAGGA GACCGTGTTC
101 ATGCAACACA GCATTATTTT ACCGCCTTTA CCCCAGCTTC TTCATACACA
151 TGCACCTTGT AAGGGCTCTT TGGCTGAAGA GAAGTTAGAA GTTCCAGAT
201 ATGGAGGGGT ATTTTCAGCA GATATGCCCA CCGCCATGGT TTTGTAGCT
251 CTGTAGGGTG GTCTTGACAC CTGCTCACTG CTGGCATCAC CTGAGCCTAT
301 GGCAGATACC CAGTGTGCTC CGCCACCATG TGAATTCATC AGCTCTGCAG
351 GCACAGACCT TGCACATAGG ATGGGCTGGG ACGCCACCCT CTGCCTCTTA
401 CCATTCACTG GGTTCGGCAA GTGTGCTGGG ATCTGGAATC ACATGGATGA
451 GGAACCCGAT AATGGTGACG ACCGAGGTAG CAGGCGAACC ACTGGCCAGG
501 GCAGGAAGTG GGCAGCTCAC GGGACTATGG CTGCACCGCG GGTTCATACC
551 GACTACCATC CTGGAGGTGG GAGCGCATGC TCATCTGTAA AAGTCCGGTC
601 CCACGTTGGA CACACCGGGG TCTTCTTCTT TGTGACCAAG GATCCTCTGG
651 CAGTGTCTTT AACAAAGCCAG AGTCTGATCC CACCGCTCAT AAAGCCAGGG
701 TTGTTGAAAG CTGGGGGCTT CCTCCTCCTC TGTGCGCAGC CCTCAGCAAA
751 CGGTACACAG CTGTGCTGTC TGCTGTACAC CGACTTGGTA TCATCCCATG
801 AACTGTCCCC CTTTCGTGCT CTGTGCTTAG GGCCTCTGA TGCCCCATCT
851 GCCTGCGCTT CTGCAACTG TTTAGCAAGC ACCTATTATC TATAGGGTGC
901 TGGGGTGCTG GCGGAGGCCA ATCGCTCCTA TTACTTTCTG CCCTGGGGAC
951 GTCTGTGTTT CCCACCTACC CCGTAACGC CTCTGCTCTG CCTTCCCATC
1001 TGCGGGGCTA ACGCCATCCC ACAAGGGCTG GGCTGTCCGT TCAGAAGAGA
1051 AACTGGGAAG GGGCCTTGAG GACCTGTGTC CAGGCAGGGT GGACAAGGGC
1101 TTTGTGCAGG GAGCTCCTCT CCCATCTTTG TGTCTGACA GCGGTGACCG
1151 TGACCCCTCA AAGCAGAGCC AGTAGTGATC AGTATCCTGC TGCTTCAAGC
1201 CTGCACGGTC CTCTTCTCCT CTCCGCACAT CTGCATGCCT GTCAAACCCA
1251 GAGTAGTTTG GGGCCTGGTA AACAGAGGGA AGTTGGCTGG AGGAGGCCAG
1301 TCAGGAGTGC AAGAACCCCG CGTACTCTGT CCCACGTGGA TAAAGTCTCT
1351 AATTCCAGTC TGAGGTGAAT TCTTAGAGAG TGCTTTCATT TAATGTTTGC
1401 TTTATGCATT TCCCCTGCAG CTGTGACTAA TTGTGGAACA GCATACATTT
1451 TGTTTTGAGA CTCTCTTGAG ATTTTCTTGG CAGTGTAAGG TCTACCCAT
1501 TTTCTCTCA GCATCAGAGA AGGCAGAAAG CAAGAGAAAG GAATGCAATG
1551 TGAGCAAGGC CAGGCACACT TGTGCTACTG CAGTTGGCAA GAATGGAGTC
1601 TAATCCCAGC ACTTTGGGAG GCCGAGGCGG GTGGATCACC TGAGGTCAGG
1651 AATTTGAGAC CAACCTGGCC AACATGTTGA AACCTCGTCT GTACTAAAAA
1701 TACAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 299 bp to 892 bp; peptide length: 198
Category: putative protein

```

1 MADTQCCPPP CEFISSAGTD LALGMGWDAT LCLLPFTGFG KCAGIWNHMD
51 EEPDNGDDRG SRRTTGQGRK WAAHGTMAAP RVHTDYHPGG GSACSSVKVR
101 SHVGHTGVFF FVDQDPLAVS LTSQSLIPPL IKPGLLKAWG FLLCAQPSA
151 NGHSLCCLLY TDLVSSHEL S PFRALCLGPS DAPSACASCN CLASTYYL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24e23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_24e23, frame 2

Report for DKFZphfkd2_24e23.2

```

{LENGTH}      198
{MW}           20948.98
{pI}           6.01
{PROSITE}      MYRISTYL      5
{PROSITE}      AMIDATION     1
{PROSITE}      CAMP_PHOSPHO_SITE 1
{PROSITE}      CK2_PHOSPHO_SITE 1
{PROSITE}      PKC_PHOSPHO_SITE 2
{KW}           All_Beta
{KW}           LOW_COMPLEXITY 6.06 %

SEQ  MADTQCCPPPCFEISSAGTDLALGMGWDATLCLLPFTGFGKCAGIWNHMDDEEPDNGDDRG
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SRRTTGQGRKWAHGTMAAPRVHTDYHPGGGSACSSVKVRSHVGHTGVFFVDQDPLAVS
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LTSQSLIPPLIKPGLLKAWGFLLCAQPSANGHSLCCLLYTDLVSSHELSPFRALCLGPS
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eccccccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccc

SEQ  DAPSACASCNCLASTYYL
SEG  .....
PRD  ccccccccccccccccccc

```

Prosites for DKFZphfkd2_24e23.2

```

PS00004      62->66  CAMP_PHOSPHO_SITE  PDOC00004
PS00005      61->64  PKC_PHOSPHO_SITE  PDOC00005
PS00005      96->99  PKC_PHOSPHO_SITE  PDOC00005
PS00006      165->169 CK2_PHOSPHO_SITE  PDOC00006
PS00008      18->24  MYRISTYL          PDOC00008
PS00008      60->66  MYRISTYL          PDOC00008
PS00008      89->95  MYRISTYL          PDOC00008
PS00008      91->97  MYRISTYL          PDOC00008
PS00008      134->140 MYRISTYL          PDOC00008
PS00009      67->71  AMIDATION         PDOC00009

```

(No Pfam data available for DKFZphfkd2_24e23.2)

DKFZphfkd2_24n20

group: intracellular transport and trafficking

DKFZphfkd2_24n20.3 encodes a novel 366 amino acid protein with similarity to human eps8 binding protein e3B1 and spectrins.

The new protein contains an Src homology domain 3 and is similar to human eps8 SH3 domain binding protein 1 (e3B1) and spectrins. Eps8 is a substrate of receptor tyrosine kinases involved in mitogenic signaling. Spectrin is part of the submembrane cytoskeletal network in the human erythrocyte ghost. Nonerythroid spectrins are proposed to have roles in cell adhesion, establishment of cell polarity, and attachment of other cytoskeletal structures to the plasma membrane. The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

strong similarity to eps8 binding protein e3B1

complete cDNA, complete cds, few EST hits
potential start at Bp 300, but there are ATGs in other frames in
5' region of the cDNA

Sequenced by GBF

Locus: /map="17"

Insert length: 1719 bp

Poly A stretch at pos. 1699, polyadenylation signal at pos. 1680

```
1  GGGGACAGCT  GCGCCGACCT  TGGCTTCCTC  TGCTGGGTGG  GATTGGGGGG
51  TGGGCCCCCA  AATGGGCCCC  TGGCTTCCCC  CTTCCTCTGG  GCAGGGGACA
101  GAGAGACACA  GGCTCGGGGA  GCAGGACTGA  CTTCCTCTTG  TCCCGGAATG
151  AGCATGCCCT  CCCTTTGCAA  GCAGGTTTGG  GTCTCAGCA  GAGGAAACCA
201  AAAGCAATAA  GAGGGAGGGA  AGGCAGAGCA  ACCAATCAAG  GGCAGGGTGA
251  GACTCAAAC  GAGCGGGCTC  CCTGGGGAGC  CAGACAGAGG  TGGGGGGTGA
301  TGGCGGAGCT  ACAGCAGCTG  CAGGAGTTTG  AGATCCCCAC  TGGCCGGGAG
351  GGTCTGAGGG  GCAACACAG  TGCCCTGCTG  CGGGTCGCTG  ACTACTGCGA
401  GGACAACTAT  GTGCAGGCCA  CAGACAAGCA  GAAGGCGCTG  GAGGAGACCA
451  TGGCCTTCAC  TACCCAGGCA  CTGGCCAGCG  TGGCCTACCA  GGTGGGCAAC
501  CTGGCCGGGC  AACTCTGCG  CATGTTGGAC  CTGCAGGGGG  CCGCCCTGCG
551  GCAGGTGGAA  GCGCGTGTA  GCACGCTGGG  CCAGATGGTG  AACATGCATA
601  TGGAGAAGGT  GGCCCGAAGG  GAGATCGGCA  CCTAGCCAC  TGTCCAGCGG
651  CTGCCCCCG  GCCAGAAGGT  CATCGCCCCA  GAGAACCTAC  CCCCTCTCAC
701  GCCCTACTGC  AGGAGACCCC  TCAACTTTGG  CTGCCTGGAC  GACATTGGCC
751  ATGGGATCAA  GGACCTCAGC  ACGCAGCTGT  CAAGAACAGG  CACCCTGTCT
801  CGAAAGAGCA  TCAAGGCCCC  TGCCACACCC  GCCTCCGCCA  CCTTGGGGAG
851  ACCGCCCCGG  ATTCCCGAGC  CAGTGCACCT  GCGGGTGGTG  CCGCAGCGCA
901  GACTCTCCGC  CGCCTCTCT  GCGTCTTCCC  TGGCCTCGGC  CGGCAGCGCC
951  GAAGGTGTCT  GTGGGGCCCC  CACGCCCAAG  GGGCAGGCAG  CACCTCCAGC
1001  CCCACCTCTC  CCCAGCTCCT  TGGACCCACC  TCCTCCACCA  GCAGCCGTCG
1051  AGGTGTTCCA  GCGGCCCTCC  ACGCTGGAGG  AGTTGTCCTC  ACCCCACCGG
1101  GACGAAGAGC  TGCCCTGCCC  ACTGGACCTG  CCTCCTCCTC  CACCCCTGGA
1151  TGGAGATGAA  TTGGGGCTGC  CTCCACCCCC  ACCAGGATTT  GGGCCTGATG
1201  AGCCCAGCTG  GGTGCCCTGC  TCATACTTGG  AGAAAGTGGT  GACACTGTAC
1251  CCATACACCA  GCCAGAAGGA  CAATGAGCTC  TCCTTCTCTG  AGGGCACTGT
1301  CATCTGTGTC  ACTCGCCGCT  ACTCCGATGG  CTGGTGCGAG  GCGTCAGCT
1351  CGGAGGGGAC  TGGATTCTTC  CCTGGGAAGT  ATGTGGAGCC  CAGCTGCTGA
1401  CAGCCCAGGG  CTCTCTGGGC  AGCTGATGTC  TGCAGTGAAG  GGGTTTCATG
1451  AGCCCCAAGC  CAAAACACGC  TCCAGTCACA  GCTGGACTGG  GTCTGCCAC
1501  CTCTTGGGCT  GTGAGCTGTG  TTCTGTCCTT  CCTCCCATCG  GAGGGAGAAG
1551  GGGTCTGGGG  GAGAGAGAAT  TTATCCAGAG  GCCTGCTGCA  GATGGGGGAG
1601  AGCTGGAAGC  CAAGAAGTTT  GTCAACAGAG  GACCCCTACT  CCATGCAGGA
1651  CAGGGTCTCC  TGCTGCAAGT  CCCAAGTTTG  AATAAACAG  ATGATGTCCA
1701  AAAAAAAAAA  AAAAAAAAAA
```

BLAST Results

Entry AC004797 from database EMBL:
Homo sapiens chromosome 17, clone hRPC.62.0.9, complete sequence.
Score = 2316, P = 5.9e-255, identities = 464/465
7 exons Bp 93317-110902

Medline entries

97163405:
Isolation and characterization of e3B1, an eps8 binding protein that regulates cell growth.

98256293:
Identification of a candidate human spectrin Src homology 3 domain-binding protein suggests a general mechanism of association of tyrosine kinases with the spectrin-based membrane skeleton.

Peptide information for frame 3

ORF from 300 bp to 1397 bp; peptide length: 366
Category: strong similarity to known protein

1 MAELQQLQEF EIPTGREALR GNHSALLRVA DYCEDNYVQA TDKQKALEET
51 MAFTTQALAS VAYQVGNLAG HTLRMLDLQG AALRQVEARV STLGMVMNMH
101 MEKVARREIG TLATVQRLPP GQKVIAPENL PPLTPYCRRP LNFGCLDDIG
151 HGIKDLSTQL SRTGTLRKS IKAPATPASA TLGRPPRIPE PVHLPVVPDG
201 RLAAASSASS LASAGSAEGV GGAPTPKGQA APPAPPLPSS LDPPPPPAV
251 EVFQRPPTLE ELSPPPPDEE LPLPLDLPPP PPLDGDELGL PPPPPGFGPD
301 EPSWVPASYL EKVVTLYPYT SQKDNELSFS EGTVICVTRR YSDGWCEGVS
351 SEGTGFPPGN YVEPSC

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_24n20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfd2_24n20, frame 3

Report for DKFZphfd2_24n20.3

[LENGTH] 366
[MW] 38947.21
[pI] 4.93
[HOMOL] TREMBL:U87166.1 gene: "SSH3BP1"; product: "spectrin SH3 domain binding protein 1"; Homo sapiens spectrin SH3 domain binding protein 1 (SSH3BP1) mRNA, complete cds. 3e-48

[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YGR136w] 9e-06
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YGR136w] 9e-06
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPR154w] 3e-05
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR388w] 2e-04
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR388w] 2e-04

[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDR162c] 4e-04
[BLOCKS] BL50002B Src homology 3 (SH3) domain proteins profile
[SUPFAM] SH3 homology 6e-17
[PROSITE] MYRISTYL 6
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 6
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 1
[PFAM] Src homology domain 3
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 24.04 %

SEQ MAELQQLQEF EIPTGREALRGNHSALLRVADYCEDNYVQATDKQKALEETMAFTTQALAS
SEG
laboA

SEQ VAYQVGNLAGHTLRMLDLQGAALRQVEARVSTLGMVMNMHMEKVARREIGTLATVQRLPP
SEG
laboA

```

SEQ      GQKVIAPENLPPLTPYCRRLNFGCLDDIGHGIKDLSTQLSRTGTLRKSIAKAPATPASA
SEG      .....
laboA    .....

SEQ      TLGRPPRIPEPVHLPVVPDGRLSAASSASSLASAGSAEGVGGAPTPKGQAAPPAPPLPSS
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
laboA    .....

SEQ      LDP PPPPAAVEVFQRPPTLEELSPPPDEELPLPLDLPPPPPLDGDDELGLPPPPPGFGPD
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
laboA    .....

SEQ      EPSWVPASYLEKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSEGTGFFPGN
SEG      xx.....
laboA    .....EECCCBCCCTTTBCCBTTEEEEEEEETTTEEEEEETTEEEEEEGG

SEQ      YVEPSC
SEG      .....
laboA    GEEE..

```

Prosites for DKFZphfkd2_24n20.3

PS00001	22->26	ASN_GLYCOSYLATION	PDOC00001
PS00004	339->343	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	72->75	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	225->228	PKC_PHOSPHO_SITE	PDOC00005
PS00005	321->324	PKC_PHOSPHO_SITE	PDOC00005
PS00005	338->341	PKC_PHOSPHO_SITE	PDOC00005
PS00006	14->18	CK2_PHOSPHO_SITE	PDOC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	321->325	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	215->221	MYRISTYL	PDOC00008
PS00008	332->338	MYRISTYL	PDOC00008

Pfam for DKFZphfkd2_24n20.3

```

HMM_NAME      Src homology domain 3

HMM            *pyVIALYDYqAqdpDELSFkEGDIIIIIEdsDD.WWrgRnnnTNGQEGW
               ++V+ LY+Y++Q ++ELSF EG +I + + D W++G + +G+
Query          311  EKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSE---GTGF      356

HMM            IPSNYVEPi*
               +P NYVEP
Query          357  FPGNYVEPS      365

```

DKFZphfkd2_24p5

group: intracellular transport and trafficking

DKFZphfkd2_24p5 encodes a novel 811 amino acid protein which is a novel splice variant of human ankyrin G.

The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments.

The new protein can find application in modulating the structure and membrane topology of Ranvier nodes and other neuronal cell membranes.

Human ankyrin G (ANK-3) new splice variant

splice variant
potential frame shift at 2720 was checked
see BLASTX

Sequenced by EMBL

Locus: /map="10q21"

Insert length: 3470 bp

Poly A stretch at pos. 3459, no polyadenylation signal found

```

1 AGCTTTAAAA GGATGCTGTC GAAGTGGTCA AAAGGATCTT AACCTCAATT
51 AAGTGGGGTT TTTTAAAAAG ATTTTGTGGG GGGCCTGAAA TTTTGAAAAT
101 CTTCGAACTC TGAGTGGGGA AAGATGTATA ATTCTCAAT TGCCTACGAG
151 GATATCAAGA TGCTGAGAGG AATTCAGCGG TGGTGAAGAG AGTGGATACA
201 AACCGAGGAT TGGTTTCCTT GAGCTGTTT GGAGGTTGAT TCTAAATCAC
251 TGCTTAAGGA ATTCCTGGAA ACATCAGGAA AACATTTGAT CATCCAAGCC
301 TAGTGGAAAT GGGTTTACCG CAGAGTGAAG ATGCAATGAC CGGGGACACA
351 GACAAATATC TTGGGCCACA GGACCTTAAG GAATTGGGTG ATGATTCCCT
401 GCCTGCAGAG GGTACATGG GCTTAGTCT CCGAGCGCGT TCTGCCAGCC
451 TCCGCTCCTT CAGTTCGGAT GGGTCTTACA CCTTGAACAG AAGCTCCTAT
501 GCACGGGACA GCATGATGAT TGAAGAACTC CTCGTGCCAT CCAAAGAGCA
551 GCATCTAACA TTCACAAGGG AATTGATTC AGATTCTCTT AGACATTACA
601 GCTGGGCTGC AGACACCTTA GACAATGTCA ATCTTGTTC AAGCCCCATT
651 CATTTCTGGT TTCTGGTTAG CTTTATGGTG GACGCGAGAG GGGGCTCCAT
701 GAGAGGAAGC CGTCATCAGC GGATGAGAAT CATCATTCCT CCACGCAAGT
751 GTACGGCCCC CACTCGAATC ACCTGCCGTT TGGTAAAGAG ACATAAACTG
801 GCCAACCACAC CCCCCATGGT GGAAGGAGAG GGATTAGCCA GTAGGCTGGT
851 AGAAATGGGT CCTGCAGGGG CACAATTTT AGGCCCTGTC ATAGTGGAAA
901 TCCTCTCACT TGGGTCCATG AGAGGAAAAG AGAGAGAACT CATTGTTCTT
951 CGAAGTGAAA ATGGTGAAAC TTGGAAGGAG CATCAGTTTG ACAGCAAAAA
1001 TGAAGATTTA ACCGAGTTAC TTAATGGCAT GGATGAAGAA CTTGATAGCC
1051 CAGAAGAGTT AGGGA AAAAG CGTATCTGCA GGATTATCAC GAAAGATTTC
1101 CCCAGTATT TTGCAGTGGT TTCCCGGATT AAGCAGGAAA GCAACCAGAT
1151 TGGTCTTGAA GGTGGAATTC TGAGCAGCAC CACAGTGCCC CTTGTTCAAG
1201 CATCTTTCCC AGAGGGTGCC CTAACATAAA GAATTCGAGT GGGCCTCCAG
1251 GCCCAGCCTG TTCCAGATGA AATTGTGAAA AAGATCCTTG GAAACAAAGC
1301 AACTTTTAGC CCAATTGTCA CTGTGGAACC AAGAAGACGG AAATTCCTATA
1351 AACCAATCAC AATGACCATT CCGGTGCCCC CGCCCTCAGG AGAAGGTGTA
1401 TCCAATGGAT ACAAAGGGGA CACTACACCC AATCTGCGTC TTCTCTGTAG
1451 CATTACAGGG GGCACCTTCG CTGCTCAGTG GGAAGACATC ACAGGAACAA
1501 CTCCTTTGAC GTTTATAAAA GATTGTGTCT CTTTACAAC CAATGTTTCA
1551 GCCAGATTTT GGCTTGCGAG CTGCCATCAA GTTTAGAAA CTGTGGGGTT
1601 AGCCACGCAA CTGTACAGAG AATTGATATG TGTTCATAT ATGGCCAAAT
1651 TTGTTGTTTT TGCCAAAATG AATGATCCCG TAGAATCTTC CTTGCGATGT
1701 TTCTGCATGA CAGATGACAA AGTGGACAAA ACTTTAGAGC AACAAGAGAA
1751 TTTTGGAGGA GTCGCAAGAA GCAAAGATAT TGAGGTTCTG GAAGGAAAAC
1801 CTATTTATGT TGATTGTTAT GGAATTTTGG CCCCACCTAC CAAAGGAGGA
1851 CAGCAACTTG TTTTAACTT TATTCTTTC AAAGAAAATA GACTGCCATT
1901 TTCCATCAAG ATTAGAGACA CCAGCCAAGA GCCCTGTGGT CGTCTGTCTT
1951 TTCTGAAAGA ACCAAAGACA ACAAAGGAC TGCCTCAAAC AGCGGTTTGC
2001 AACTTAAATA TCACTCTGCC AGCACATAAA AAGATTGAGA AAACAGATGG
2051 ACCACAGAGC TTCGCATCCT TAGCTTTACG TAAGCGCTAC AGCTACTTGA
2101 CTGAGCCTGG AATGAGTCCA CAGAGTCCAT GTGAACGGAC AGATATCAGG
2151 ATGGCAATAG TAGCCGATCA CCTGGGACTT AGTTGGACAG AACTGGCAAG
2201 GGAACCTGAAT TTTTCAGTGG ATGAAATCAA TCAAATACGT GTGGAAAATC
2251 CAAATCTCTT AATTTCTCAG AGCTTCATGT TTTTAAAAAA ATGGGTACC
2301 AGAGACGGAA AAAATGCCAC AACTGATGCC TTAACCTCGG TCTTGACAAA
2351 AATTAATCGA ATAGATATAG TGACACTGCT AGAAGGACCA ATATTTGATT

```

```

2401 ATGGAATAT TTCAGGCACC AGAAGTTTTC CAGATGAGAA CAATGTTTTC
2451 CATGACCTTG TTGATGGTTA TCCTTCCCTT CAAAGTGAAC TGGAAACCCC
2501 CACAGGGTTG CACTACACAC CACCTACCCC TTTCCAGCAA GATGATTATT
2551 TTAGTGATAT CTCTAGCATA GAATCTCCCC TTAGAACCCC TAGTAGACTG
2601 AGTGATGGGC TAGTGCCTTC CCAGGGGAAC ATAGAGCATT CCGCAGATGG
2651 ACCTCCAGTC GTAAGTGCAG AAGACGCTTC CTTAGAAGAC AGCAAACTGG
2701 AAGACTCAGT GCCTTTAACA GAAATGCCTG AAGCAGTGAT GTAGATGAGA
2751 GCCAGTTGGA GAATGTATGT CTGAGTTGGC AGAATGAGAC ATCAAGTGGG
2801 AACCTAGAGT CCTGCGCTCA AGCTCGAAGA GTAAGTGGTG GGTACTAGA
2851 TCGACTGGAT GACAGCCCTG ACCAGTGTAG AGATTCCATT ACCTCATATC
2901 TCAAAGGAGA AGCTGGCAAA TTTGAAGCAA ATGGAAGCCA TACAGAAATC
2951 ACTCCAGAAG CAAAGACAAA ATCTTACTTT CCAGAATCCC AAAATGATGT
3001 AGGAAAACAG AGTACCAAGG AAACCTCTGAA ACCAAAAATA CATGGATCTG
3051 GTCATGTTGA AGAACCAGCA TCACCACTAG CAGCATATCA GAAATCTCTA
3101 GAAGAAACCA GCAAGCTTAT AATAGAAGAG ACTAAACCTT GTGTGCCTGT
3151 CAGTATGAAA AAGATGAGTA GGACTTCTCC AGCAGATGGC AAGCCAAGGC
3201 TTAGCCTCCA TGAAGAAGAG GGGTCCAGTG GGTCTGAGCA AAAGCAGGGA
3251 GAAGGTTTTT AGGTGAAAAC GAAGAAAGAA ATCCGGCATG TGGAAAAGAA
3301 GAGCCACTCG TAACAGCGAA CGGTCACTCA AGGATCATAA GTTTTTACTG
3351 CAGTATTGTA GAAATTCGTG GAAGAAATGT CAGCAGGAAG TAAAAATCA
3401 CCGAGAAGTG TGTGTGTGTT CGCTGCTTCC ACACATTAAT GGCATGATT
3451 TTTTATGCA AAAAAAAAAA

```

BLAST Results

Entry MMANK3A_1 from database TREMBL:
 Ank3"; product: "ankyrin 3"; Mus mu... +3 4022 0.0 2

Entry HS13616 from database EMBL:
 Human ankyrin G (ANK-3) mRNA, complete cds.
 Length = 14,770
 Plus Strand HSPs:
 Score = 8505 (1276.1 bits), Expect = 0.0, Sum P(3) = 0.0
 Identities = 1799/1873 (96%)

Medline entries

95394457:
 Chromosomal localization of the ankyrinG gene
 (ANK3/Ank3) to human 10q21 and mouse 10.

95138209:
 A new ankyrin gene with neural-specific isoforms localized at the
 axonal initial segment and node of Ranvier

Peptide information for frame 3

ORF from 309 bp to 2741 bp; peptide length: 811
 Category: known protein
 Classification: unset

```

1 MALPQSEDAM TGDTDKYLGP QDLKELGDDS LPAEGYMGFS LGARSASLRS
51 FSSDGSYTLN RSSYARDSMM IEELLVPSKE QHLTFTREFD SDSLRHYSWA
101 ADTLDNVNLV PSPIHSGFLV SFMVDARGGS MRQSRHHGMR IIIPPRKCTA
151 PTRITCRVLK RHKLANPPPM VEGEGLASRL VEMGPAGAQF LGPVIVEIPH
201 FGSMRGKERE LIVLRSENGE TWKEHQFDSK NEDLTELLNG MDEELDSPEE
251 LGKKRICRII TKDFPQYFAV VSRIKQESNQ IGPEGGILSS TTVPLVQASF
301 PEGALTKRIR VGLQAQVPVD EIVKKILGNK ATFSPIVTV EPRRRKFHKPI
351 TMTIPVPPPS GEGVSNKYKG DTTPNLRLLC SITGGTSPAQ WEDITGTTPL
401 TFKKDCVSFT TNVSARFWLA DCHQVLETVG LATQLYRELI CVPYMAKFVV
451 FAKMNDPVES SLRCFCMTDD KVDKTLEQQE NFEEVARSKD IEVLEGKPIY
501 VDCYGNLAPL TKGGQQLVFN FYSFKENRLP FSIKIRDSQ EPCGRLSFLK
551 EPKTTKGLPQ TAVCNLNLITL PAHKKIEKTD GRQSFASLAL RKRYSYLTP
601 GMSPPQSPER TDIRMAIVAD HLGLSWTELA RELNFSVDEI NQIRVENPNS
651 LISQSFMLK KVVTRDGKNA TTDALTSVLT KINRIDIVTL LEGPIFDYGN
701 ISGTRSFAD E NNVFHDPVDC YPSLQVELET PTGLHYTPPT PFOQDDYFSD
751 ISSIESPLRT PSRLSDGLVP SQGNIEHSAD GPPVVTAEDA SLEDSKLEDS
801 VPLTEPFAV M

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24p5, frame 3

TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds., N = 1, Score = 4022, P = 0

TREMBL:MMANK3B_3 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

TREMBL:MMANK3B_4 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

>TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds.
Length = 1,094

HSPs:

Score = 4022 (603.5 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 769/805 (95%), Positives = 783/805 (97%)

```
Query:      1 MALPQSEDAMTGDTDKYLGPQDLKELGDDSLPAEGYMGFSLGARSASLRSFSSDGSYTLN 60
             MALP SEDA+TGDTDKYLGPQDLKELGDDSLPAEGY+GFSLGARSASLRSFSSD SYTLN
Sbjct:      1 MALPHSEDAITGDTDKYLGPQDLKELGDDSLPAEGYVGFSLGARSASLRSFSSDRSYTLN 60

Query:      61 RSSYARDSMMIEELLVPSKEQHLTFTRFSDSLRHYSWAADTLDNVNLVPSPIHSGFLV 120
             RSSYARDSMMIEELLVPSKEQHLTFTRFSDSLRHYSWAADTLDNVNLV SP+HSGFLV
Sbjct:      61 RSSYARDSMMIEELLVPSKEQHLTFTRFSDSLRHYSWAADTLDNVNLVSSPVHSGFLV 120

Query:      121 SFMVDARGGSMRGRHGMRIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180
             SFMVDARGGSMRGRHGMRIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL
Sbjct:      121 SFMVDARGGSMRGRHGMRIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180

Query:      181 VEMGPAGAQLGFPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLTELLNG 240
             VEMGPAGAQLGFPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDL ELLNG
Sbjct:      181 VEMGPAGAQLGFPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLAELLNG 240

Query:      241 MDEELDSPEELGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300
             MDEELDSPEELG KRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF
Sbjct:      241 MDEELDSPEELGTRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300

Query:      301 PEGALTKRIRVGLQAQPVPEIVKKILGNKATFSPIVTEPRRRKFHKPITMTIPVPPPS 360
             PEGALTKRIRVGLQAQPVPEIVKKILGNKATFSPIVTEPRRRKFHKPITMTIPVPPPS
Sbjct:      301 PEGALTKRIRVGLQAQPVPEETVKKILGNKATFSPIVTEPRRRKFHKPITMTIPVPPPS 360

Query:      361 GEGVSNYKGDTPNLRLLCSITGCTSPAQWEDITGTPLTFIKDCVSFTTNVSARFWLA 420
             GEGVSNYKGD TPNLRLLCSITGCTSPAQWEDITGTPLTFIKDCVSFTTNVSARFWLA
Sbjct:      361 GEGVSNYKGDATPNLRLLCSITGCTSPAQWEDITGTPLTFIKDCVSFTTNVSARFWLA 420

Query:      421 DCHQVLETVGLATQLYRELICVPYMAKFVVFAMNDPVESSLRCFCMTDDKVDKTLQQE 480
             DCHQVLETVGLA+QLYRELICVPYMAKFVVFAM NDPVESSLRCFCMTDD+VDKTLQQE
Sbjct:      421 DCHQVLETVGLASQLYRELICVPYMAKFVVFAMTNDPVESSLRCFCMTDDRVDKTLQQE 480

Query:      481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSENRLPFSIKIRDTSQ 540
             NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSENRLPFSIKIRDTSQ
Sbjct:      481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSENRLPFSIKIRDTSQ 540

Query:      541 EPCGRLSFLKEPKTKGLPQTAVCNLNLITLPAHKKIEKTDGRQSFASLALRKRYSYLTEP 600
             EPCGRLSFLKEPKTKGLPQTAVCNLNLITLPAHKK EK D RQSFASLALRKRYSYLTEP
Sbjct:      541 EPCGRLSFLKEPKTKGLPQTAVCNLNLITLPAHKKA EKADRRQSFASLALRKRYSYLTEP 600

Query:      601 GMSQSPCERDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSFMLK 660
             MSPQSPCERDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSFMLK
Sbjct:      601 SMSQSPCERDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSFMLK 660

Query:      661 KKWTRDGKNATTDALTSVLTAKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDVPDG 720
             KKWTRDGKNATTDALTSVLTAKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDVPDG
Sbjct:      661 KKWTRDGKNATTDALTSVLTAKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDVPDG 720

Query:      721 YPSLQVELETPTGLHYTPPTPFQDDYFSDISSIESPLRTPSRLSDGLVPSQGNIEHSAD 780
             +PS QVELETPT GL++TPP PFQDD+FSDISSIESP RTPSRLSDGLVPSQGNIEH
Sbjct:      721 HPSFQVELETPTGLHYWTPPNPFQDDHFSDISSIESPFRTPSRLSDGLVPSQGNIEHPTG 780

Query:      781 GPPVVTAEDASLEDSKLEDSVPLTE 805
             GPPVVTAED SLEDSK++DSV +T+
```

Sbjct: 781 GPPVVTAEDTSLEDSKMDDSVTVTD 805

Pedant information for DKFZphfkd2_24p5, frame 3

Report for DKFZphfkd2_24p5.3

[LENGTH] 811
 [MW] 90104.66
 [pI] 5.40
 [HOMOL] TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial
 ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds. 0.0
 [BLOCKS] BL50017B Death domain proteins profile
 [PIRKW] phosphoprotein 0.0
 [PIRKW] alternative splicing 0.0
 [PIRKW] peripheral membrane protein 0.0
 [PIRKW] cytoskeleton 0.0
 [SUPFAM] ankyrin 0.0
 [SUPFAM] ankyrin repeat homology 0.0
 [SUPFAM] unassigned ankyrin repeat proteins 0.0
 [KW] TRANSMEMBRANE 2
 [KW] LOW_COMPLEXITY 1.73 %

SEQ MALPQSEDAMTGDTDKYLGFPQDLKELGDDSLPAEGYMGFSLGARSASLSRFSSDGSYTLN
 SEG
 PRD ccc
 MEM

SEQ RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVPSPIHSGFLV
 SEG
 PRD cccchhhhhhhhhheeehhhhhhhhhhcccccccccccccccccccccccccccccccccc
 MEMMMMMMMMMMMMM

SEQ SFMVDARGGSMRGRHHGMRIIPPRKCTAPTRITCRILVKRHKLANPPPMVEGEGLASRL
 SEGxxxxxxxxxxxxx.....
 PRD eeeeecc
 MEM MMMMMMMMMMMMMMMM.....M

SEQ VEMGPAGAQLGPGVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLTELLNG
 SEG
 PRD ecc
 MEM MMM

SEQ MDEELDSPeelGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF
 SEG
 PRD cccccchhhhhhhhhheeecc
 MEM

SEQ PEGALTKRIRVGLQAQVPDEIVKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS
 SEG
 PRD ccchhhhhhhhhhhhhcc
 MEM

SEQ GEGVSNGYKGDTPNLRLLCSITGGTSPAQWEDITGTPTLTFIKDCVSFTTNVSARFWLA
 SEG
 PRD ccc
 MEM

SEQ DCHQVLETVGLATQLYRELICVPYMAKFVFAKMNPDVESSLRCFCMTDDKVDKTLQQE
 SEG
 PRD cchhhhhhhhhhhhhhhhhhhhhcchhhhhheeeccccchhhhhhhhhccccchhhhhhhhh
 MEM

SEQ NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFKNRPLPFSIKIRDTSQ
 SEG
 PRD ccc
 MEM

SEQ EPCGRSLFLKEPKTKGLPQTAVCNLNLITLPAHKKIEKTDGRQSFASLALRKRYSLTEP
 SEG
 PRD ccc
 MEM

SEQ GMSPOQPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNSLISQSFMLK
 SEG
 PRD cccccchhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhccceeeccccchhhhhhhhh
 MEM

```
SEQ  KVVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDVPVG
SEG  .....
PRD  hhhhhccccccchhhhhhhhhccceeeeecccccccccccccccccccccccccccccc
MEM  .....

SEQ  YPSLQVELETPTGLHYTPPTPFQDDYFSDISSIESPLRTPSRLSDGLVPSQGNIEHSAD
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  GPPVVTAEDASLEDSKLEDSVPLTEMPEAVM
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccc
MEM  .....
```

(No Prosite data available for DKFZphfkd2_24p5.3)

(No Pfam data available for DKFZphfkd2_24p5.3)

DKFZphfkd2_3113

group: transmembrane protein

DKFZphfkd2_3113 encodes a novel 406 amino acid protein with C. elegans cosmid Y37D8A and A. thaliana H71412 hypothetical protein.

The novel protein contains 3 transmembrane regions.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to A.thaliana and C.elegans;
membrane regions: 3

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="17"

Insert length: 2052 bp

Poly A stretch at pos. 2032, no polyadenylation signal found

```

1 AGTGACGTGA GCGGGTTCCG GTTGTCTGGA GCCAGCGGC GGGTGTGAGA
51 GTCGGTAAGG AGCAGCTTCC AGGATCCTGA GATCCGGAGC AGCCGGGGTC
101 GGAGCGGCTC CTCAGAGTT ACTGATCTAT GAAATGGCAG AGAATGGAAA
151 AAATTGTGAC CAGAGACGTG TAGCAATGAA CAAGGAACAT CATAATGGAA
201 ATTTACACAGA CCCCTCTTCA GTGAATGAAA AGAAGAGGAG GGAGCGGGAA
251 GAAAGGCAGA ATATTGTCTT GTGGAGACAG CCGCTCATT CTTGTCAGTA
301 TTTTCTCTG GAAATCCTTG TAATCTTGAA GGAATGGACC TCAAAATTAT
351 GGCATCGTCA AAGCATTTGT GTGTCTTTT TACTGCTGCT TGCTGTGCTT
401 ATAGCTACGT ATTATGTTGA AGGAGTGCAT CAACAGTATG TGCAACGTAT
451 AGAGAAACAG TTTCTTTTGT ATGCCTACTG GATAGGCTTA GGAATTTTGT
501 CTTCTGTTGG GCTTGGAAAC GGGCTGCACA CCTTCTGCT TTATCTGGGT
551 CCACATATAG CCTCAGTTAC ATTAGCTGCT TATGAATGCA ATTCAGTTAA
601 TTTTCCCGAA CCACCTATC CTGATCAGAT TATTTGTCCA GATGAAGAGG
651 GCACTGAAGG AACCATTTT TTGTGGAGTA TCATCTCAAA AGTTAGGATT
701 GAAGCCTGCA TGTGGGGTAT CGGTACAGCA ATCGGAGAGC TGCCTCCATA
751 TTTTCATGGC AGAGCAGCTC GCCTCTCAGG TGCTGAACCA GATGATGAAG
801 AGTATCAGGA ATTTGAAGAG ATGCTGGAAC ATGCAGAGTC TGCACAAGAC
851 TTTGCCTCCC GGGCCAAACT GGCAGTTCAA AAAGTAGTAC AGAAAGTTGG
901 ATTTTGTGGA ATTTTGGCCT GTGCTTCAAT TCCAAATCCT TTATTGTATC
951 TGGCTGGAAT AACGTGTGGA CACTTCTG GACCTTTTGT GACCTTCTTT
1001 GGTGCAACCC TAATTGGAAA AGCAATAATA AAAATGCATA TCCAGAAAAT
1051 TTTTGTATATA ATAACATTCA GCAAGCACAT AGTGGAGCAA ATGGTGGCTT
1101 TCATTGGTGC TGTCCCGGC ATAGGTCCAT CTCTGCAGAA GCCATTTTCA
1151 GAGTACCTGG AGGCTCAACG GCAGAAGCTT CACCACAAAA GCGAAATGGG
1201 CACACCAAGG GGAGAAAAC GGTGTCTCTG GATGTTTGAA AAGTGTGGTC
1251 TTGTCATGGT GTGTTACTTC ATCCTATCTA TCATTAATC CATGGCACA
1301 AGTTATGCCA AACGAATCCA GCAGCGGTTG AACTCAGAGG AGAAAACTAA
1351 ATAAGTAGAG AAAGTTTAA ACTGCAGAAA TTGGAGTGG TGGGTCTGTC
1401 CTTAAATTGG GAGGACTCCA AGCCGGGAAG GAAATTTCCC TTTTCCAACC
1451 TGTATCAATT TTTACAACCT TTTTCTGAA AGCAGTTTAG TCCATACTTT
1501 GCACTGACAT ACTTTTCTCT TCTGTGCTAA GGTAAAGGTAT CCACCTCGA
1551 TGCAATCCAC CTTGTGTTT CTTAGGGTGG AATGTGATGT TCAGCAGCAA
1601 ACTTGCAACA GACTGGCCTT CTGTTTGTTA CTTTCAAAAG GCCACATGA
1651 TACAATTAGA GAATTCACAC CGCACAAAAA AAGTTCTTAA GTATGTTAAA
1701 TATGTCAAGC TTTTTAGGCT TGTCAAAAT GATTGCTTTG TTTTCTAAG
1751 TCATCAAAAT GTATATAAAT TATCTAGATT GGATAACAGT CTTGCATGTT
1801 TATCATGTTA CAATTTAATA TTCCATCTG CCCAACCTT CCTCTCCCAT
1851 CCTCAAAAAA GGGCCATTTT ATGATGCATT GCACACCTC TGGGGAAATT
1901 GATCTTTAAA TTTTGAGACA GTATAAGGAA AATCTGGTTG GTGCTTTACA
1951 AGTGAGCTGA CACCATTTT TATTCTGTGT ATTTAGGATG AAGTCTTGAA
2001 AAAAATTCTA TAAAGACATC TTTAATCATT CCAAAAAAAA AAAAAAAA
2051 AA

```

BLAST Results

Entry AC004686 from database EMBL:
 *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 17, clone
 hRPC.1073 F_15; HTGS phase 1, 8 unordered pieces.
 Score = 4142, P = 6.1e-199, identities = 830/832

Medline entries

~ No Medline entry

Peptide information for frame 2

ORF from 134 bp to 1351 bp; peptide length: 406
Category: similarity to unknown protein

```

1 MAENGKNCQ RRVAMNKEHH NGNFTDPSSV NEKKRREREE RQNVILWRQP
51 LITLQYFSLE ILVILKEWTS KLWHRQSIVV SFLLLLAVLI ATYYVEGVHQ
101 QYVQRIEKQF LLYAYWIGLG ILSSVGLGTG LHTFLLYLGP HIASVTLAAY
151 ECNSVNFPEP PYPDQIICPD EEGTEGTIFL WSIISKVRIE ACMWGIGTAI
201 GELPPYFMAR AARLSGAEPD DEEYQEFEEH LEHAESAQDF ASRAKLAVQK
251 LVQKVGFFGI LACASIPNPL FDLAGITCGH FLVPFWTFFG ATLIGKAIK
301 MHIQKIFVII TFSKHIVEQM VAFIGAVPGI GPSLQKPFQE YLEAQRQKLH
351 HKSEMGTPOG ENWLSWMFEK LVVVMVCYFI LSIINSMAQS YAKRIQQLN
401 SEEKTK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_3i13, frame 2

TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; *Caenorhabditis elegans* cosmid Y37D8A, N = 1, Score = 905, P = 8.8e-91

TREMBL:ATAC98_2 gene: "YUP8H12.2"; *Arabidopsis thaliana* chromosome 1 YAC yUP8H12 complete sequence., N = 1, Score = 470, P = 1.1e-44

PIR:H71412 hypothetical protein - *Arabidopsis thaliana*, N = 1, Score = 293, P = 6e-24

>TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; *Caenorhabditis elegans* cosmid Y37D8A

Length = 457

HSPs:

Score = 905 (135.8 bits), Expect = 8.8e-91, P = 8.8e-91
Identities = 167/317 (52%), Positives = 228/317 (71%)

```

Query:   38 REERQNVILWRQPLITLQYFSLEILVILKEWTSKLWHRQSIVVSFLLLLAVLIATYYVEG 97
          R ER+ IV WR+P I + Y +EI + E K+ +++++ + + + + Y+ G
Sbjct:   93 RMERETIVFWRRPHIVIPYALMEIAHLAVELFFKILAHKTVLLLTAISIGLAVYGYHAPG 152

Query:   98 VHQQYVQRIEKQFLLYAYWIGLGILSSVGLGTGLHTFLLYLGPFIASVTLAAYECNSVNF 157
          HQ++VQ IEK L +++W+ LG+LSS+GLG+GLHTFL+YLGPHIA+VT+AAYEC S++F
Sbjct:  153 AHQEHVQTIEKHILWWSWVLLGVLSIGLGSGLHTFLIYLGPHIAAVTMAAYECQSLDF 212

Query:  158 PEPYPYDQIICPDDEEGTEGTIFLWSIISKVRIEACMWGIGTAIGELPPYFMARAARLSGA 217
          P+PPYP+ I CP + + F W I++KVR+E+ +WG GTA+GELPPYFMARAAR+SG
Sbjct:  213 PQQPYPESIQCPSTKSSIAVTF+WQIVAKVRVESLLWGAGTALGELPPYFMARAARISGQ 271

Query:  218 EPDDEEYQEFEEHLE-HAESAQD----FASRAKLAVQKLVQKVGFFGILACASIPNPLFD 272
          EPDDEEY+EF E++ ES D RAK V+ + ++GF GIL ASIPNPLFD
Sbjct:  272 EPDDEEYREFLELMNADKESDADQKLSIVERAKSWVEHNIHRLGFPFGILLFASIPNPLFD 331

Query:  273 LAGITCGHFLVPFWTFFGATLIGKAIKMHQKIFVIITFSKHIVEQMVAFIGAVPGIGP 332
          LAGITCGHFLVPFW+FFGATLIGKA++KMH+Q FVI+ FS H E V + +P +GP
Sbjct:  332 LAGITCGHFLVPFWFFGATLIGKALVKMHVQMGEVILAFSDHHAENFVKILEKIPAVGP 391

Query:  333 SLQKPFQEYLEAQRQKLH 350
          +++P + LE QR+ LH
Sbjct:  392 YIRQPIISDLLEKQRKALH 409

```

Pedant information for DKFZphfd2_3i13, frame 2

Report for DKFZphfd2_3i13.2

```

SEQ      MAENGKNCDDRRVAMNKEHHNGNFTDPSSVNEKKRREREERQNIVLWRQPLITLQYFSLG
SEG      .....XXXXXXXXXX.....
PRD      cccccccchhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhccccchhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      ILVILKEWTSKLWHRQSIVSVFLLLLAVLIATYYVEGVHQYVQRIEKQFLLYAYWIGLG
SEG      .....XXXXX.....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhhhhhhhh
MEM      MM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      ILSSVGLGTGLTFTLLYLGPFIASVTLAAYECNSVNFPEPPYPDQICPDEEGTEGTIFL
SEG      XXXXXXXXXXXXX.....
PRD      hccccccccceeeeeeeccchhhhhhhhhhhhhccccccccccccccccccccccccceeee
MEM      .....

SEQ      WSIISKVRIEACMWGIGTAIGELPPYFMAARAARLSGAEPDDEEYQEFEEMLEHAESAQDF
SEG      .....XXXXXXXXXXXXXXXXXX.....
PRD      eehhhhhhhhhhhhhhhccccccccccccchhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      ASRAKLAVQKLVQKVGFFGILACASIPNPLFDLAGITCGHFLVPFWTFFGATLIGKAIK
SEG      .....
PRD      hhhhhhhhhhhhhhhhhccccccccccccccccccccccccceeeehhhhhhhhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      MHIQKIFVITFSKHIVEQMVAFIGAVPGIGPSLQKPFQYEYLAQRQKLHHKSEMGTPOG
SEG      .....
PRD      hhhhheeeeeeechhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccc
MEM      .....

SEQ      ENWLSWMFEKLVVVMVCYFILSIINSMASQSYAKRIQORLNSEEKTK
SEG      .....
PRD      cchhhhhhhhhhhheeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc
MEM      .....

```

PS00001	23->27	ASN_GLYCOSYLATION	PDOC00001
PS00005	69->72	PKC_PHOSPHO_SITE	PDOC00005
PS00006	29->33	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	126->132	MYRISTYL	PDOC00008
PS00008	173->179	MYRISTYL	PDOC00008
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	197->203	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	275->281	MYRISTYL	PDOC00008
PS00008	325->331	MYRISTYL	PDOC00008
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	356->362	MYRISTYL	PDOC00008

399

DKF2phfkd2_3ol7

group: metabolism

DKF2phfkd2_3ol7 encodes a novel 72 amino acid protein with similarity to bos taurus NADH-ubiquinone oxidoreductase B33 subunit (EC 1.6.5.3) (EC 1.6.99.3).

NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. The novel protein is the human orthologue of bovine B22.

The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

strong similarity to bovine NADH-UBIQUINONE OXIDOREDUCTASE B22 subunit

complete cDNA, complete cds, EST hits,
in frame stop codon at ~274 will be checked
ESTs HS1291620/AA883920 show no stop codon at this side

Sequenced by BMFZ

Locus: unknown

Insert length: 693 bp
Poly A stretch at pos. 670, polyadenylation signal at pos. 659

```

1 CAGCAGGCGT GCAGTTTCCC GGCTCTCCGC GCGGCCGGGG AAGGTCAGCG
51 CCGTAATGGC GTTCTTGGCG TCGGGACCCT ACCTGACCCA TCAGCAAAAG
101 GTGTTGCGGC TTTATAAGCG GCGCTACGC CACCTCGAGT CGTGGTGCGT
151 CCAGAGAGAC AAATACCGAT ACTTTGCTTG TTTGATGAGA GCCCGGTTTG
201 AAGAACATAA GAATGAAAAG GATATGGCGA AGGCCACCCA GCTGCTGAAG
251 GAGGCCGAGG AAGAATTCTG GTAACGTCAG CATCCACAGC CATACATCTT
301 CCCTGACTCT CCTGGGGGCA CCTCCTATGA GAGATACGAT TGCTACAAGG
351 TCCCAAGATG GTGCTTAGAT GACTGGCATC CTCTGAGAA GGCAATGTAT
401 CCTGATTACT TTGCCAAGAG AGAACAGTGG AAGAACTGC GGAGGGAAAG
451 CTGGGAACGA GAGGTTAAGC AGCTGCAGGA GGAAACGCCA CCTGGTGGTC
501 CTTTAACTGA AGCTTTGCCC CTGCCCCGAA AGGAAGGTGA TTTGCCCCCA
551 CTGTGGTGGT ATATTGTGAC CAGACCCCGG GAGCGGCCCA TGTAGAAAGA
601 GAGAGACCTC ATCTTTCATG CTTGCAAGTG AAATATGTGA CAGAACATGC
651 ACTTGCCCTA ATAAAAAATC AGTAAAAAAA AAAAAAATAA AAA

```

BLAST Results

Entry S28256 from database PIR:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
>TREMBL:MIBTCIB22_1 gene: "cI-B22"; product: "NADH-ubiquinone
oxidoreductase complex B22 subunit"; B.taurus mitochondrion cI-B22
mRNA for B22 subunit of the NADH-ubiquinone oxidoreductase complex
Score = 933, P = 5.2e-93, identities = 163/179, positives = 172/179,
frame +2

Medline entries

92389317
Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from RT bovine heart mitochondria.
Application of a novel strategy for RT sequencing proteins using the polymerase chain reaction

Peptide information for frame 2

ORF from 56 bp to 271 bp; peptide length: 72
Category: strong similarity to known protein

```

1 MAFLASGPYL THQQKVLRLY KRALRHLESW CVQRDKYRYF ACLMRARFEE
51 HKNEKDMAKA TQLLKEAEEE FW*ROHPQPY IFPDSPGGTS YERYDCYKVP
101 EWCLDDWHPs EKAMYPDYFA KREQWKKLRR ESWEREVKQL QEETPPGGPL
151 TEALPPARKE GDLPLLWYI VTRPRERPM

```

BLASTP hits

Sequences producing significant alignments: (bits) Value

sp|Q02369|NI2M_BOVIN|OD36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE... 141 7e-34
tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO NADH-UBIQ... 53 3e-07

>sp|Q02369|NI2M_BOVIN|OD36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE
OXIDOREDUCTASE B22 SUBUNIT (EC 1.6.5.3) (EC 1.6.99.3)
(COMPLEX I-B22) (CI-B22).[BOS TAURUS]
Length = 178

Score = 141 bits (351), Expect = 7e-34
Identities = 63/71 (88%), Positives = 68/71 (95%)

Query: 2 AFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKAT 61
AFL+SG YLTHQQKVLRLYKRALRHLESWC+ RDKYRYFACL+RARF+EHKNEKDM KAT
Sbjct: 1 AFLSSGAYLTHQQKVLRLYKRALRHLESWCIIHRDKYRYFACLLRARFDEHKNEKDMVKAT 60

Query: 62 QLLKEAEEEFW 72
QLL+EAEFEFW
Sbjct: 61 QLLREAEEEFW 71

>tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO
NADH-UBIQUINONE OXIDOREDUCTASE B22.[CAENORHABDITIS
ELEGANS]
Length = 163

Score = 52.7 bits (124), Expect = 3e-07
Identities = 25/64 (39%), Positives = 41/64 (64%), Gaps = 1/64 (1%)

Query: 10 LTHQQKVLRLYKRALRHLESWCVQRD-KYRYFACLMRARFEEHKNEKDMAKATQLLKEAE 68
L+H+QKV RLYKR LR +++W + + R+ C++RARF+ + +E D K+ LL +
Sbjct: 12 LSHRQKVTRLYKRCLEVDNWWYGGNNLEVRFQKCIIRARFDANAEVDTRKSQILLADGC 71

Query: 69 EEFW 72
+ W
Sbjct: 72 RQLW 75

Alert BLASTP hits for DKFZphfkd2_3o17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_3o17, frame 2

Report for DKFZphfkd2_3o17.2

[LENGTH] 72
[MW] 8839.28
[pI] 9.26
[HOMOL] PIR:S28256 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
2e-34
[KW] All_Alpha

SEQ MAFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKA
PRD cccccccchhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhcchhhhhhh

SEQ TQLLKEAEEEFW
PRD hhhhhhhhhccc

(No Prosite data available for DKFZphfkd2_3o17.2)

(No Pfam data available for DKFZphfkd2_3o17.2)

DKFZphfkd2_46a6

group: kidney derived

DKFZphfkd2_46a6 encodes a novel 315 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="228.6 cR from top of Chr15 linkage group"

Insert length: 2774 bp

Poly A stretch at pos. 2751, polyadenylation signal at pos. 2732

```

1 CTGCGAGCG CAGCTATGGC TGCTGGCGTA CCCTGTGCGT TAGTCAACCAG
51 CTGCTCCTCC GTCTTCTCAG GAGACCAGCT GGTCCAACAT ACCCTTGGAA
101 CAGAAAGATCT TATTGTGGAA GTGACTTCCA ATGATGCTGT GAGATTTTAT
151 CCCTGGACCA TTGATAATAA ATACTATTCA GCAGACATCA ATCTATGTGT
201 GGTGCCAAAC AAATTTCTTG TTAAGTCAGA GATTGCAGAA TCTGTCCAAG
251 CATTTGTGGT TTAAGTTGAC AGCACACGAA AATCGGGCCT TGATAGTGTC
301 TCCTCATGGC TTCCACTGGC AAAAGCATGG TTACCTGAGG TGATGATCTT
351 GGTCTGCGAT AGAGTGTCTG AAGATGGTAT AAACCGACAA AAAGCTCAAG
401 AATGGAGCCT CAAACATGGC TTTGAATTGG TAGAACTTAG TCCAGAGGAG
451 TTGCCTGAGG AGGATGATGA CTTCCCAGAA TCTACAGGAG TAAAGCGAAT
501 TGTCCAAGCC CTGAATGCCA ATGTGTGGTC CAATGTAGTG ATGAAGAATG
551 ATAGGAACCA AGGCTTTAGC CTTCTCAACT CATTGACTGG AACAAACCAT
601 AGCATTGGGT CAGCAGATCC CTGTCACCCA GAGCAACCCC ATTTGCCAGC
651 AGCAGATAGT ACTGAATCCC TCTCTGATCA TCGGGGTGGT GCATCTAACA
701 CAACAGATGC CCAGGTTGAT AGCATTGTGG ATCCCATGTT AGATCTGGAT
751 ATTCAAGAAT TAGCCAGTCT TACCACCTGA GGAGGAGATG TGGAGAATTT
801 TGAAGAGACC TTTTCAAAGT TAAAGGAAAT GAAAGACAAG GCTCCGACGC
851 TTCTCATGTA GCAAAGAAAA GTGCATGCAG AAAAGGTGGC CAAAGCATTG
901 TGGATGGCAA TCGGGGGAGA CAGAGATGAA ATTGAAGGCC TTTTCATCTGA
951 TGGAGAGCAC TGAATTATTC ATACTAGGGT TTGACCAACA AAGATGCTAG
1001 CTGTCTCTGA GATACCTCTC TACTCAGCCC AGTCATATTT TGCCAAATTT
1051 GCCCTTATCA TGTTGGCTGC CTGACTTGTT TATAGGGTCC CCTTAATTTT
1101 AGTTTTTAGT AGGAGGTTAA GGAGAAATCT TTTTTTTCCT CAGTATATTG
1151 TAAGACAGTG AGGAATACAG TGATAGTAAT GAGTGAGGAT TTCTTAAATA
1201 TACTTTTTTT TTGTTCTAGG AATGAGGGTA GGATAAATCT CAGAGGCTCTG
1251 TGATGATTTA TCAAGTTGAA GACAACCTCC AGGCCATTCC TGGTCAACCT
1301 TTTAAGTAGC ATTTCCAGCA TTCACACTTG ATACTGCACA TCAGGAGTTG
1351 TGTCACCTTT CCTGGGTGAT TTGGGTTTTT TCCATTCAAG GAGCTTGTAG
1401 CTCTGAGCTA TGATGCTTTT ATTGGGAGGA AAGGAGGCAG CTCGAGAATT
1451 GATGTGAGCT ATGTGGGGCC GAAGTCTCAG CCCGCAGCTA AGTCTCTACC
1501 TAAGAAATG CCTCTGGGCA TTCTTTTGAA GTATAGTGTC TGAGCTCATG
1551 CTAGAAAGAA TCAAAAAGCC AGTGTGGATT TTTAGGCTGT AATAAATGAG
1601 GCAAAAGGAT TCTATTCCAG TGGGAAGGAA ACCTCTCTAC TGAGTTGTGG
1651 GGGATATGTT GTATGTTAGA GAGAACCTTA AGGAGTCCTT GTATGGGCCA
1701 TGGAGACAGT ATGTGATAAC ATACCGTGAT TTTTCATGAAG AAATCTTCTT
1751 GTCCCTAGAG TCTCCCCTGC TGCTTGAGAT GCCAGAGCTG TGTGTTGCA
1801 CACCTGCAAA ACAAGGCACA TTTCCCCTTT TCTCTTTAAA GCCAAAGAGA
1851 GATCACTGCC AAAGTGGGAG CACTAAGGGG TGGGTGGGGA AGTGAAATGT
1901 TAGGCGATGA ATTCCTGAGC ACCTTGTTTT TCTTCCAAGG TTCGTAGCTC
1951 CTCTCTGCCC TTCCAAGCCT GTAACCTCGG AGGACTATCT TTTGTTCTCT
2001 ATCCTTTGTC TTGTTAGAGT GGGTCAGCCC CAGAGGAAGT GATAAGCAAA
2051 TGGCAAGTTT TTAAGGAAG AGTGGAAAGT ACTGCAATAA AAAATCCCTTA
2101 TTTGTTTTTG TAGACTTTGT AATGCATATC ATTAGCCCTC ACTGTGATCA
2151 TTAAGTCTGT GGCTCTGAAC TGGCAGATAG TACAGTGGAT GGAAGGTGCC
2201 CGCACACCAAG CTGAGAACTG GTTCTGGCCT AGGTGGGCTC TAGAACCATT
2251 TACACAGCAT GAAAGAAACA GGTGGGTTA GGAGCAGAAA GAAATAAGGC
2301 TCACACCCCT CCAGACACTA CCTTATAAGC ACTGCAGAAC CTGAAACAGA
2351 TGGCAGAAGG AATGGAATGC TACAGGGGCC AGCAGGAGTG ACCACAGGGA
2401 GGGGACAGCT CAGTGACTGG AGCATTACAG AAGAGGCTTT CCAGGGAACA
2451 CTGGACATTG CTTAGTGACC TTTTGTTCCT TTTTTTTTTT TTTTCTTTTA
2501 CTGTTTCTGAA AGACTTTGAG TCTGTGGTTC ACCACCAGCC CATCAGTGTT
2551 TCTTTGAGGT GATTGCATTA GGAAGTTGG CTCTGGGATT GCAAAAAAAA
2601 AAAAAAGGTG GAACATGTTT TCCTTAAAG ATGGAAGGTT TTAGAAAATA
2651 TACTAGGCCA TCTGGTTAGA AAAAAAGAC CAGACTAGAA AAAGCTGTGA

```

2701 ATTTGATTTT GTAGATTAAA CAAAGCCAGA TGATTAAAT GTGATTATT
2751 TATAAAAAA AAAAAAAA AAAA

BLAST Results

Entry HS463358 from database EMBL:
human STS WI-14364.
Length = 472
Minus Strand HSPs:
Score = 1605 (240.8 bits), Expect = 5.0e-68, P = 5.0e-68
Identities = 347/361 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 16 bp to 960 bp; peptide length: 315
Category: putative protein
Classification: unset

1 MAAGVPCALV TSCSSVFSGD QLVQHTLGTE DLIVEVTSND AVRFPWTID
51 NKYYADINL CVVPNKFLVT AEIAESVQAF VVYFDSTRKS GLDSVSSWLP
101 LAKAWLPEVM ILVCDRVSED GINRQKAQEW SLKHGFELVE LSPEELPEED
151 DDFESTGVK RIVQALNANV WSNVVMKNDR NQGFSLNLSL TGTNHSIGSA
201 DPCHEQPHL PAADSTESLS DHRGGASNTT DAQVDSIVDP MLDLDIQELA
251 SLTTGGGDVE NFERPFSKLK EMKDKAATLP HEQRKVHAEK VAKAFWMAIG
301 GDRDEIEGLS SDGEH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46a6, frame 1

PIR:T04362 probable GTP-binding protein yptm3 - maize, N = 1, Score =
87, P = 0.21

PIR:S71585 GTP-binding protein GB2 - Arabidopsis thaliana, N = 1, Score
= 86, P = 0.27

>PIR:T04362 probable GTP-binding protein yptm3 - maize
Length = 210

HSPs:

Score = 87 (13.1 bits), Expect = 2.4e-01, P = 2.1e-01
Identities = 34/160 (21%), Positives = 67/160 (41%)

Query: 48 TIDNKYYADINLCVVPNKFL-VTAEIAESVQAFVVYFDSTRKSGLDSVSSWLPLAKAWL 106
TIDNK I F +T ++ +D TR+ + ++SWL A+
Sbjct: 49 TIDNKFIKLQIWDTAGQESFRSITRSYYRGAAGALLVYDITRRETFNHLASWLEDARQHA 108
Query: 107 PE---VMIL--VCDRVSEDGINRQKAQEWSLKHGFELVELSPEELPEEDDDFFESTGVKR 161
VM++ CD ++ ++ +++++HG +E S + ++ F ++ G
Sbjct: 109 NANMTVMLIGNKCDLSHRRVSYEEGEQFAKEHGLVFMEASAKTAQNVEEAFIKTAGT-- 166
Query: 162 IVQALNANVWSNVVMKNDRNQGFSLNLSLTGTNHSIGSADPC 203
I + + ++ N G+++ NS G S A C
Sbjct: 167 IYKKIQDGIFDVSNESNGIKVGYAVPNSSGGGAGSSSQAGGC 208

Pedant information for DKFZphfkd2_46a6, frame 1

Report for DKFZphfkd2_46a6.1

[LENGTH] 315

[MW] 34505.54
[pI] 4.55
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 6.67 %

SEQ MAAGVPCALVTSCSSVFSGDQLVQHTLCTEDLIVEVTSNDAVRFPWTIDNKYY SADINL
SEG
PRD cccccceeeccccccccceeeccccceeeccccceeeccccccccccccce

SEQ CVVPNKFLVTAIEASVQAFVVFYDSTRKSGLDVSSWLPLAKAWLPEVMILVCDRVSED
SEG
PRD eeccccchhhhhhhhhheeeccccccccccccccccccccceeeccccccc

SEQ GINRQKAQEWSLKHGFELVELSPEELPEEDDDFPESTGVKRIVQALNANVMSNVVMKNDR
SEGxxxxxxxxxxxxxxxxxxxxx.....
PRD cchhhhhhhhhccceeeccccccccccccccccchhhhhhhccceeecccc

SEQ NQGFSLLNSLTGTNHSIGSADPCHPEQPHLPAADSTESLSDHRGGASNTTDAQVDSIVDP
SEG
PRD cch

SEQ MLDLDIQELASLTGGGDVENFERPFSKLEMKDKAATLPHEQRKVHAEKVAKAFWMAIG
SEG
PRD hhhhhhhhhhhccceccccccccchhhhhhhhhhhccchhhhhhhhhhhhhhhc

SEQ GDRDEIEGLSSDGEH
SEG
PRD ccccccccccccc

(No Prosite data available for DKFZphfd2_46a6.1)

(No Pfam data available for DKFZphfd2_46a6.1)

DKFZphfkd2_46b10

group: kidney derived

DKFZphfkd2_46b10.1 encodes a novel 315 amino acid protein with similarity to C.elegans cosmid F25B5.3

The novel protein contains a HTH-LYSR-family PROSITE pattern. Proteins of the lysR family are bacterial transcriptional regulatory proteins which bind DNA using a helix-turn-helix motif. Most of these proteins are transcription activators and usually negatively regulate their own expression. They all possess a potential 'helix-turn-helix' DNA-binding motif in their N-terminal section. The 'helix-turn-helix' motif is missing in DKFZphfkd2_46a6.1. No informative BLAST results, no predictive PFAM or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to C.elegans F25B5.3

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 1285 bp

Poly A stretch at pos. 1266, no polyadenylation signal found

```
1 CAGTCTACGC GAGCTGCCTG TTTTTCCT GCTGGACGC GCATGAGGGC
51 CCCGTCCATG GACCGCGCGG CCGTGGCGAG GGTGGGCGCG GTAGCGAGCG
101 CCAGCGTGTG CGCCCTGGTG GCGGGGGTGG TGCTGGCTCA GTACATATTC
151 ACCTTGAAGA GGAAGACGGG GCGGAAGACC AAGATCATCG AGATGATGCC
201 AGAATCCAG AAAAGTTCAG TTCGAATCAA GAACCTACA AGAGTAGAAG
251 AAATTATCTG TGGTCTTATC AAAGGAGGAG CTGCCAACT TCAGATAATA
301 ACGGACTTTG ATATGACACT CAGTAGATTT TCATATAAAG GGAAAAGATG
351 CCCAACATGT CATAATATCA TTGACAACTG TAAGCTGGTT ACGGATGAAT
401 GTAGAAAAAA GTTATTGCAA CTAAAGGAAA AATATTACGC TATTGAAGTT
451 GATCCTGTTC TTACTGTAGA AGAGAAGTAC CCTTATATGG TGGAAATGGTA
501 TACTAAATCA CATGGTTTGC TTGTTCAGCA AGCTTTACCA AAAGCTAAAC
551 TTAAAGAAAT TGTGGCAGAA TCTGACGTTA TGCTCAAAGA AGGATATGAG
601 AATTTCTTTG ATAAGCTCCA ACAACATAGC ATCCCCGTGT TCATATTTTC
651 GGCTGGAATC GCGGATGTAC TAGAGGAAGT TATTCGTCAA GCTGGTGTTC
701 ATCATCCCAA TGTCAAAGTT GTGTCCAATT TTATGGATTT TGATGAAACT
751 GGGGTGCTCA AAGGATTTAA AGGAGAACTA ATTCATGTAT TTAACAAACA
801 TGATGGTGCC TTGAGGAATA CAGAAATATT CAATCAACTA AAAGACAATA
851 GTAACATAAT TCTTCTGGGA GACTCCCAAG GAGACTAAG AATGGCAGAT
901 GGAGTGGCCA ATGTTGAGCA CATCTGAAA ATTGGATATC TAAATGATAG
951 AGTGGATGAG CTTTATAGAA AGTACATGGA CTCTTATGAT ATTGTTTTAG
1001 TACAAGATGA ATCATTAGAA GTAGCCAACCT CTATTTTACA GAAGATTCTA
1051 TAAACAGACA TTCTCCAAGA AGACCTCTCT CCTGTGGGTG CAATTGAACT
1101 GTTCATCCGT TCATCTTGCT GAGAGACTTA TTTATAATAT ATCCTTACTC
1151 TCGAAGTGTT CCCTTTGTAT AACTGAAGTA TTTTCAGATA TGGTGAATGC
1201 ATTGACTGGA AGCTCCTTTT CTCCACCTCT CTCAACACAC TCCTCACCGT
1251 ATCTTTTAAC CCATTTAAAA AAAAAAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 43 bp to 1050 bp; peptide length: 336

Category: similarity to unknown protein

Classification: unset

Prosite motifs: HTH_LYSR_FAMILY (16-47)

```

1 MRAPSMDDRAA VARVGAVASA SVCALVAGVV LAQYIFTLKR KTGRKTKIIE
51 MMPEFQKSSV RIKNPTRVEE IICGLIKGGA AKLQIITDFD MTLRSFSYKG
101 KRCPTCHNII DNCKLVTDEC RKKLLQLKEK YYAIEVDPVL TVEEKYPYMV
151 EWYTKSHGLL VQALPKAKL KEIVAESDVM LKEGYENFFD KLOQHSIPVF
201 IFSAGIGDVL EEVIRQAGVY HPNVKVVSNF MDFDETVGLK GFKGELIHVF
251 NKHDGALRNT EYFNQLKDNS NIILLGDSQG DLRMADGVAN VEHLKIGYL
301 NDRVDELLEK YMDSYDIVLV QDESLEVANS ILQKIL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_46b10, frame 1

SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III., N = 1, Score = 524, P = 2.2e-50

TREMBL:AC005499_12 gene: "T6A23.12"; Arabidopsis thaliana chromosome II BAC T6A23 genomic sequence, complete sequence., N = 2, Score = 194, P = 1.4e-26

>SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.

Length = 376

HSPs:

Score = 524 (78.6 bits), Expect = 2.2e-50, P = 2.2e-50
Identities = 112/300 (37%), Positives = 174/300 (58%)

```

Query:   44 RKTIIEMMPEFQ--KSSVRIKNPTRVEEIIICGLIKGGA AKLQIITDFD MTLRSFSYK-G 100
          +KT ++ ++ + + + + +PT V + ++ GGA K +I+DFD TLSRE+ + G
Sbjct:   73 KKTQVPLLMNYLLGEEQILVADPTAVA AKLRKVMVGGAGKTVVISDFD YTLRSFANEQG 132

Query:  101 KRCPTCHNIID-NCKLVTDECRKLLQLKEKYAIEVDPVLTVEEKYPYMV EWYTKSHGL 159
          +R T H + D N + E +K + LK KYI E P LT+EEK P+M +W+ SH L
Sbjct:  133 ERLSTTHGVFDDNVMRLKPELGQKRFVDLKNKYPIEFSPNLTME EKIPHMEKWWGTS HSL 192

Query:  160 LVQALPKAKLKEIVAESDVM LKEGYENFFD KLOQHSIPV FIFSAGIGDVL EEVIRQA-G 218
          +V + K +++ V +S ++ K+G E+F + L H+IP+ IFSAGIG+++E ++Q G
Sbjct:  193 IVNEKFSKNTIEDFVRQSRIVFKDGAEDFIEALDAHNIPLVIFSAGIGNIIEYFLQQKLG 252

Query:  219 VYHPNVKVVS NFMD FDETVGLK GFKGELIHVFNKHDGAL-RNTEYFNQLKDNS NIILLGD 277
          N +SN + FDE F LIH F K+ + + T +F+ + N+ILLGD
Sbjct:  253 AIPRNTHFISNMILFDEDDNACAFSEPLIHTFCKNSSVIQKETSFFHDIAGRVNVILLGD 312

Query:  278 SQGDLRMADGVANVEHLKIGYLNDRVDEL--LEKYMDSYDIVLVQDESLEVANS ILQKI 335
          S GD+ M GV LK+GY N +D+ L+ Y + YDIVL+ D +L VA I+ I
Sbjct:  313 SMGDIHMDVGVERDGP TLKVGYYNGSLDDTAALQHYEEVYDIVLIHDPTLNVAQKIVDII 372

```

Pedant information for DKFZphfd2_46b10, frame 1

Report for DKFZphfd2_46b10.1

```

[LENGTH]      336
[MW]           37948.37
[pI]           6.67
[HOMOL]        SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.
3e-51
[PROSITE]      HTH_LYSR_FAMILY 1
[KW]           TRANSMEMBRANE 2
[KW]           LOW_COMPLEXITY 7.44 %

```

```

SEQ  MRAPSMDDRAA VARVGAVASASVCALVAGVVLAQYIFTLKRKTGRKTKIEMMPEFQKSSV
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD  cccchhhhhccchhhhhheeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeehhhhhhhhhe
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  RIKNPTRVEEIIICGLIKGGA AKLQIITDFD MTLRSFSYKGRKRCPTCHNIIDNCKLVTDEC
SEG  .....
PRD  eeccccchhhhhhhhhhhcccccceeecccccceeecccccceeecccccceeecccccchhhhhh
MEM  .....

```

```
SEQ  RKKLLQLKEKYAIEVDPVLTVEEKYPYMVEWYTKSHGLLVQQALPKAKLKEIVAESDVM
SEG  .....
PRD  hhhhhhhhhhhheeeccccccccchhhhhccccchhhhhccccchhhhhhhhhhhcc
MEM  .....

SEQ  LKEGYENFFDKLQHSIPVFIFSAGIGDVLEEVIHQAGVYHPNVKVVSFMDFDGTGLK
SEG  .....
PRD  cccccchhhhhhhccccceeeccccchhhhhhhhhccccceeeccccccccce
MEM  .....MMMMMMMMMMMMMMMM.....

SEQ  GFKGELIHVFNKHDGALRNTEYFNQLKDNSNIILLGDSQGDLRMADGVANVEHILKIGYL
SEG  .....
PRD  eccccceeeccccccccchhhhhhhceeeccccccccccccccccccccceeeec
MEM  .....

SEQ  NDRVDELLEKYMDSYDIVLVQDESLEVANSILQKIL
SEG  .....
PRD  cchhhhhhhhhhhheeeecchhhhhhhhhccc
MEM  .....
```

Prosite for DKFZphfd2_46b10.1

PS00044 16->47 HTH_LYSR_FAMILY PDOC00043

(No Pfam data available for DKFZphfd2_46b10.1)

DKF2phfkd2_46d13

group: kidney derived

DKF2phfkd2_46d13 encodes a novel 506 amino acid protein with weak similarity to KE03 protein

The novel protein contains a RGD site.

No informative BLAST results; No predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KE03 protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="227.6 cR from top of Chr1 linkage group"

Insert length: 3346 bp

Poly A stretch at pos. 3328, polyadenylation signal at pos. 3308

```
1 CTCTCGCGAG AGGAGCAAGA GGAAGATGGC CGTGCCCTGT TTTTCGGTGT
51 AAGGCAGCAG ACGGCGGCTG CGACGGCGAG ACTGAGATCC TGGTGTCTGT
101 GGCACCTGAG TTCTAGCTTC CCCAGCGAG CGCGCGTCCC TTCGTGCCTA
151 GGCAGAGGCC GGCTCTTCCC CGGGAGATGC GTTGTGCCCA GGTCTGGGGG
201 CTCAGTGGGA GTTCATGCTG CGCTGGAGGC TCTTGGCCAC CGCTCTAATC
251 GCCTTGTGCC GCCGCGAGCG CAGCTCCGTC GCCAGCGGTG AGCCTCCCGA
301 TTCCCCCCTT TGCCCCCTGGC GCGGGCGATG ACCGGGGAGA AGATCCGCTC
351 ACTGCGGAGG GACCACAAGC CCAGCAAAGA AGAAGGGGAC CTGCTGGAGC
401 CCGGGGATGA AGAAGCGGCG GCTGCCCTCG GCGGTACCTT TACCAGAAGC
451 AGGATTGGCA AGGGCGGCAA AGCTTGTCTA AAGATCTTCA GTAACCATCA
501 CCACCGGCTA CAGCTGAAGG CAGCTCCGGC CTCCTCCAAT CCCCCGCGCG
551 CCGCGGCTCT GCCGCTGCAC AATTCCTCCG TGAAGTCCAA CTCCCACTCC
601 CCGGCCCTTC TGGCCCGCAC CAACCCCGTT GCTGTCGTCT CGGATGGAGG
651 CAGTTGCCCC GCACACTACC CGGTGCACGA GTGCGTCTTC AAGGGGGATG
701 TGAGGAGACT CTCCTCTCTC ATCCGACGCG ACAATATCGG GCAGAAAGAT
751 AATCACGGAA ATACTCCTTT ACACCTTGCT GTGATGTTAG GAAATAAAGT
801 TACAGCTCTT TTGAGGAAGC TTAAGCAGCA ATCCAGGGAA AGTGTGAAG
851 AAAAAAGCACC TCGATTATTA AAAGCCCTGA AAGAGCTAGG TGACTTTTAT
901 CTAGAAGTTC ACTGGGATTT TCAAAGCTGG GTGCCCTTAC TTTCCCGAAT
951 TCTGCCTTCC GATGCATGTA AAATATACAA ACAAGGTATC AATATCAGGC
1001 TTGACACAAC TCTCATAGAC TTTACTGACA TGAAGTGCCA ACGAGGGGAT
1051 CTAAGCTTCA TTTTCAATGG GGATGCGGCG CCCTCTGAAT CTTTGTAGT
1101 ATTAGACAAT GAACAAAAG TTTATCAGCG AATACATCAT GAGGAATCAG
1151 AGATGGAAAC AGAAGAAGAG GTGGATATTT TAATGAGCAG TGATATTTAC
1201 TCTGCAACTT TATCAACAAA ATCAATTCTT TTCACGCGTG CCGAGACAGG
1251 ATGGCTTTTT CGGGAAGATA AAACAGAAAG AGTAGGAAAC TTTTGGCAG
1301 ACTTTTACCT GGTGAATGGA CTTGTTATAG AATCAAGGAA AAGAAGAGAA
1351 CATCTCAGTG AAGAGGATAT TCTTCGAAAT AAGGCCATCA TGGAGAGTTT
1401 GAGTAAAGGT GGAACATAA TGGAACAGAA TTTTGAGCCG ATTCGAAGAC
1451 AGTCTCTTAC ACCGCCTCCT CAGAACACTA TTACATGGGA AGAATATATA
1501 TCTGCTGAAA ATGGAAAAGC TCCTCATCTG GGTAGAGAAAT TGGTGTGCAA
1551 AGAGAGTAAG AAAACGTTTA AAGCTACGAT AGCCATGAGC CAGGAATTTT
1601 CCTTAGGGAT AGAGTTATTA TTGAATGTTT TAGAAGTAGT AGCTCCCTTC
1651 AAGCACTTTA ACAAGCTTAG AGAATTGTTT CAGATGAAGC TTCCTCCAGG
1701 CTTTCTGTGA AAATTAGATA TACCTGTGTT TCCCACAATC ACAGCCACTG
1751 TGACTTTTCA GGAGTTTCGA TACGATGAAT TTGATGGCTC CATCTTTACT
1801 ATACCTGATG ACTACAAGGA AGACCCAAGC CGTTTCTCTG ATCTTTAACT
1851 GACGTGGAAA AGGATGCCGT CTAACCAAGG AAAGAAAATA CAGAGACCCT
1901 AGAAGTGGAT CCAAATAGAA GGGACAAATG CTTTCAGTGA AGAAAAGGGA
1951 ATTACACATT GAATCGACAC ATCAGTAATA CGATACAGTG AAATGGGCCT
2001 CTAATAAGAA TTTCAGCGAG TTTTCTGATG TGCCATTTTT TGTCTTTTTA
2051 AAAATATACA TATTATAAAT GTAATAGTTT GACACATTAA TGACCCTAAG
2101 ACCTGCGTAT GTGAAGCAGC TATGAGTGCT GTGATTTGTT TTTAAAAATT
2151 TTTACACTTC TTGTTGAAAT ATATATGCAT ATAAATATAT CTATATCTAT
2201 ATCTATATCT AAAACACTCC TGGACCATTA ACGTAAATTA AATGTCTTAA
2251 GAGATATGGA GCCCTTTTAA ACTTGTATC TTTATGCAAG GTGACATTTA
2301 TAAATATTCC TTCGAGCTTT GTTTTCATAA AATGTAAACT ATGTAACATT
2351 ATGTATAGTT CAGTAATTTG AATGTTTGT TCAATATAATG AACTAGAAGG
2401 AATGCAATTT TCTGTAGATG AATGAACCAA ATGGTAACCA TTAACCAATT
2451 GCATTTATAT GTTGCAATAC ATTTCAAGAG GAGCGTTCAC TCTGCAGGGA
2501 ATAAGGTACC TCCTTTAGCA CCTTAGTGCA ATTCATTGTG GTGCTATTTG
2551 TTTTACCTG AATGTTTGT ACTAATCTTC CTTTCATAGA ACCTCTATTT
2601 TTTTTTTTTT TAAACTTGAG TTTGAGTCCT TGTATGGTC ATCATAAGGT
```

```

2651 AATGGTTAGC ATGTTTAAAG ATATTCCTCT TCCAAATCTC AGCACTTTAA
2701 AAAAAAATCC AAATTTTAA ACTTGCTTCC TAATAAGTAC ACATCGGTCT
2751 GATTATTTTG TTTGTTTTTA GTAGAATATG GATGCATTGG TGTCAGTTT
2801 AAAAAACAAT ACACATATTT TGGACAACCC TACATATTTA ATCCTTTCAA
2851 AATAAGATAA AAACATTTTA TATGCTAACA GAATATATTT GTTACAAGTT
2901 AAAGTCCAGA AGTATACACA AGATTGATTA CTCCTATTAT TTTTTTAA
2951 TCACAGGAAA ATATTGATTT CATTGTCTCC AAAGTGATAA AATCTTGAT
3001 TACTCATTTT TGCACCTAAA ATTTTCTTA TTTATCCAA GGTGGTTGA
3051 AGGTCCAAGT ATGAAAATAA ATTAGGGGGA TTAATGTATA ACAGTTATA
3101 AGTATCATGT TGTATTAAAG AGCTTACTTA GATTGATGTT TTTAAATGT
3151 ATCCTGATGA ATGCTCAAG AATGCATCTG TCAAGTTTTT TAGACTGACC
3201 AGTAGCTTAA ACTTTTTTCA GGATTTTAGG TAATTTGAAA GGAGTTAGA
3251 GACCCTTATT GAAAATATGA TTTAAAAATC CAAAGCATAA ACCGTAAGAA
3301 AAATTTTAAA TAAACATCTT TAAAGCTGAA AAAAAAAAAA AAAAAA

```

BLAST Results

Entry HS121353 from database EMBL:

human STS WI-14729.

Score = 1697, P = 1.9e-69, identities = 363/379

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 1845 bp; peptide length: 506

Category: similarity to unknown protein

```

1 MTGEKIRSLR RDHKPSKEEG DLLEPGDEEA AAALGGTFTR SRIGKGGKAC
51 HKIFSNHHHR LQLKAAPASS NPPGAPALPL HNSSVTANSQ SPALLAGTNP
101 VAVVADGGSC PAHYPVHECV FKGDVRLSL LIRTHNIGQK DNHGNTPLHL
151 AVMLGNKVTA LRRKLKQQR ESVEEKRPRL LKALKELGDF YLELHWDFQS
201 WVPLLSRIPL SDACKIYKQG INIRLDTTLI DFTDMKCQRG DLSFIFNGDA
251 APSESEFVLD NEQKVYQRIH HESEMETEE EVDILMSSDI YSATLSTKSI
301 SFTRAQTGWL FREDKTERVG NFLADFYLVN GLVIESRKRR EHLSEEDILR
351 NKAIMESLSK GGNIMEQNFE PIRRQSLTPP PQNTITWEEY ISAENGKAPH
401 LGRELVCCKES KKTFKATIAM SQEFPGLIEL LLNVLEVVP FKHFNKLREF
451 VQMKLPFGFP VKLDIPVFPT ITATVTFQEF RYDEFDGSIF TIPDDYKEDP
501 SRFPDL

```

BLASTP hits

Entry CEC01F1_3 from database TREMBL:

gene: "C01F1.6"; Caenorhabditis elegans cosmid C01F1.

Score = 371, P = 4.5e-61, identities = 69/138, positives = 96/138

Entry CEC18F10_9 from database TREMBL:

gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10.

Score = 383, P = 3.4e-39, identities = 103/349, positives = 182/349

Entry AF064604_1 from database TREMBL:

product: "KE03 protein"; Homo sapiens KE03 protein mRNA, partial cds.

Score = 348, P = 8.3e-32, identities = 95/295, positives = 148/295

Alert BLASTP hits for DKFZphkd2_46d13, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphkd2_46d13, frame 1

Report for DKFZphkd2_46d13.1

```

[LENGTH]      506
[MW]           57003.12
[pI]           6.40

```



```

[OMOL]      TREMBL:CEC18F10_9 gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10. 2e-35

[BLOCKS]    BL01288E
[PROSITE]   RGD      1
[PROSITE]   MYRISTYL  7
[PROSITE]   CAMP_PHOSPHO_SITE  2
[PROSITE]   CK2_PHOSPHO_SITE   9
[PROSITE]   PKC_PHOSPHO_SITE   6
[PROSITE]   ASN_GLYCOSYLATION  1
[KW]        Alpha_Beta
[KW]        LOW_COMPLEXITY      7.51 %

SEQ  MTGEKIRSLRRDNHKSKEEGDLLEPGDEEAAAALGGTFTSRIGKGGKACHKIFSNHHHR
SEG  .....XXXXXXXXXXXXX.....
PRD  cceeeccccccccccccccccchhhhhccccccccccccceeeecchhhh

SEQ  LQLKAAPASSNPPGAPALPLHNSVSTANSQSPALLAGTNPVAVVADGGSCPAHYPVHECV
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhhccccccccceeeccccccccccccceeeccccceeeccccccccceee

SEQ  FKGDVRRLLSSLIIRTHNIGQKDNHGNTPLHLAVMLGNKVTALLRKLKQQSRESVEEKRPRL
SEG  .....
PRD  ecccchhhhhhhhhccccccccccccceeecccchhhhhhhhhhhccchhhhhhhhh

SEQ  LKALKELGDFYLELHWFQSWVPLLSRILPSDACKIYKQGINIRLDTTLIDFTDMKQCRG
SEG  .....
PRD  hhhhhccccceehhhhhccccceccccccccceeeccccceeecccccccccc

SEQ  DLSFIFNGDAAPSESFVLDNEQKVYQRIHHEESEMETEEVDILMSSDIYSATLSTKSI
SEG  .....XXXXXXXXXXXXX.....
PRD  ceeeeeccccceeeccccceehhhhhhhhhhhhhhhhhhhccccceeecccccc

SEQ  SFTRAQTGWLFRDQKTERVGNFLADFYLVNGLVIESRKRREHLSEEDILRNKAIMESLSK
SEG  .....
PRD  eeeccccceccccchhhhhheeeeeecccccccccccccccccccccccccccccccc

SEQ  GGNIMEQNFEPIRRQSLTPPPQNTITWEEYISAENGKAPHLGRELVCESKKTFKATIAM
SEG  .....
PRD  cceeeccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhh

SEQ  SQEFLGLIELLLNVLEVVAFFKHFKNKLEFVQMKLPPGFVFKLDIPVFPPTITATVTFQEF
SEG  .....
PRD  hhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccceeeeeecccccccccc

SEQ  RYDEFDGSIFTIPDDYKEDPSRFPDL
SEG  .....
PRD  ccccccccccecccccccccccccc

```

Prosite for DKFZphkd2_46d13.1

PS00001	82->86	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	373->377	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	8->11	PKC_PHOSPHO_SITE	PDOC00005
PS00005	296->299	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	336->339	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	278->282	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	386->390	CK2_PHOSPHO_SITE	PDOC00006
PS00006	476->480	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	46->52	MYRISTYL	PDOC00008
PS00008	108->114	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00016	239->242	RGD	PDOC00016

(No Pfam data available for DKFZphkd2_46d13.1)

DKFZphfkd2_46j20

group: metabolism

DKFZphfkd2_346j20 encodes a novel 224 amino acid protein similar to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein seems to be the human ortholog of 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein can find application in modulating the homoprotocatechuate degradative pathway and as a enzyme for biotechnologic production processes.

strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase

complete cDNA, complete cds, EST hits,
potential start at Bp 16 matches kozak consensus ANCatgG
strong similarity to proteins of worm plant archea and bacteria
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase is part of
the tyrosine metabolism (degradation of tyrosine late step) EC 5.3.1.-
complete cds according to similar C.elegans and A.thaliana protein

Sequenced by MediGenomix

Locus: unknown

Insert length: 1706 bp

Poly A stretch at pos. 1686, polyadenylation signal at pos. 1667

```
1  CACTTGATGG GAATCATGGC AGCATCCAGG CCATTGTCCC GCTTCTGGGA
51  GTGGGGGAAAG AACATCGTCT GCGTGGGGAG GAACTACGCG GACCAACGTC
101 GGGAGATGCG CAGCGCGGTG TTGAGCGAGC CCGTGCTGTT CCTGAAGCCG
151 TCCACGGCCT ACGCGCCCGA GGGCTCGCCC ATCCTCATGC CCGCGTACAC
201 TCGCAACCTG CACCACGAGC TGGAGCTGGG CGTGGTGATG GGCAAGCGCT
251 GCCGCGCAGT CCCCAGGGCT GCGGCCATGG ACTACGTGGG CGGCTATGCC
301 CTGTGCCTGG ATATGACCGC CCGGGACGTG CAGGACGAGT GCAAGAAGAA
351 GGGGCTGCCC TGGACTCTGG CGAAGAGCTT CACGGCGTCC TGCCCGGTCA
401 GCGCGTTTCG GCCCAAGGAG AAGATCCCTG ACCCTCACAA GCTGAAGCTC
451 TGGCTCAAGG TCAACGGCGA ACTCAGACAG GAGGGTGAGA CATCCTCCAT
501 GATTTTTTTC ATCCCTTACA TCATCAGCTA TGTTTCTAAG ATCATAACCT
551 TGAAGAAGG AGATATTATC TTGACTGGGA CGCCAAAGGG AGTTGGACCG
601 GTTAAAGAAA ACGATGAGAT CGAGGCTGGC ATACACGGGC TGGTCAGTAT
651 GACATTTTAA GTGGAAGAGC CAGAATATTG AGTTATTCTT TAACAAGTTT
701 CGAGAGAGAA GGGAGCAAGA CAAGAGCAAG CAACGGCTAT TAAATGTCAC
751 AATCCTTTAA TTAGAAACCA TTTATTGGCC GGACGCGGTG GCTCAGCCT
801 GTAATCGCAG CACTTTGGGA GGCCGAGGCG GCGGCTCAC GACGTCAGGA
851 GATCCAGACC ATCTTGGCTA ACAGGGTGAA ACCCGCTCTC TACTAAAAAT
901 ACAAAAAATT AGCCGGGCGT GGTGGCGGCG GCCTGTAGTC CCAGCTACTC
951 TGGAGGCTGA GGCAGGAGAA TCAATTGAAC CCGGAGGCGG GAGCTTACAG
1001 TGAGCTGAGA TTGCGCCACT GTACTCCTGG GCAACAGCGA GACTCCGCTC
1051 CAAAAAATAA AAAAAAATAA AGAAACCATT TATTTTAAAA ATGATTAGAT
1101 TGCTATGCCT CAACTCATAG AAGATGAACC CTTCAAGAAA ACGTGAAGTA
1151 GAACGGGTGG GCCAGAAATG AAAACAGGCA AGTAAAGTAT TTCTTCGGAA
1201 AACATTTTAT CAAACCAAT GTTAAAAAGA CTTTCTTTT GTAAACTGG
1251 ATTAGAGAAG ACTTTTCAGT GGGTTATCTC TAGGATGATC AGTAGTTCAG
1301 CACTTAAAAA CTGCAGAGAA AACTGAAAGT TATGTTCCAG ATAACCTTCC
1351 GTTGTTTACC AAATTTTCTT AGATTGGTTC ATCATCAGGA AGCATTGTGA
1401 AAAATAAAAA TCTCCACAAA TTACTGGCCC ATCTCGGACT TGCTGAATCA
1451 ATTTGATAGG ATTAATCTCC AGTGAAGCTG TGTTTACAGG GCATTCCAAG
1501 TGATTCCTTAT CAGGAAATGT GAAAAACACT CCTGTACATA ATCGGTTAAT
1551 TTAATAATTT ACTTAATAAG TGAACAAGTA ATGAAGATTT CACCTGTTTA
1601 CTTAGGGTAT CTACCCAGAC CCATCGATTG TGAGTTCGGG AGATGATTTT
1651 GAAATTACTG TTTTCCAAAT AAAGGTGCTC CCTTCCAAAA AAAAAAAAAA
1701 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

94039092: Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of *Escherichia coli* C.

Peptide information for frame 1

ORF from 7 bp to 678 bp; peptide length: 224
Category: strong similarity to known protein

1 MGIMAASRPL SRFWEWGKNI VCVGRNYADH VREMRSVLS EPVFLFKPST
51 AYAPEGSPIL MPAYTRNLHH ELELGVVMGK RCRAVPEAAA MDYVGGYALC
101 LDMTARDVQD ECKKKGLPWT LAKSFTASCP VSAFVPKEKI PDPHKLKLWL
151 KVNGLRQEG ETSSMIFSIP YIISYVSKII TLEEGDIILT GTPKGVGPVK
201 ENDEIEAGIH GLVSMTEKVE KPEY

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfk2_46j20, frame 1

PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans*, N = 1, Score = 537, P = 8.7e-52

PIR:D71109 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase - *Pyrococcus horikoshii*, N = 1, Score = 529, P = 6.1e-51

PIR:C71425 hypothetical protein - *Arabidopsis thaliana*, N = 1, Score = 519, P = 7e-50

PIR:A64864 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase b1180 - *Escherichia coli*, N = 1, Score = 474, P = 4.1e-45

>PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans*
Length = 214

HSPs:

Score = 537 (80.6 bits), Expect = 8.7e-52, P = 8.7e-52
Identities = 99/211 (46%), Positives = 138/211 (65%)

Query: 10 LSRFEWVGKNI VCVGRNYADH VREMRSVLS EPVFLFKPST AYAPEGSPIL MPAYTRNLH 69
L+ F IVCVGRNY DH E+ +A+ +P+LF+K ++ EG PI+ P +NLH
Sbjct: 4 LAGFRNLATKIVCVGRNYKDHALELGNAIPKKPMLFVKT VNSFIVEGEPIVAPPGCQNLH 63

Query: 70 HELELGVVMGKRCRAVPEAAA MDYVGGYALCLDMTARDVQDECKKKGLPWT LAKSFTASC 129
E+ELGVV+ K+ + ++ AMDY+GGY + LDMTARD QDE KK G PW LAKSF SC
Sbjct: 64 QEVELGVVISKASRISKSDAMDYIGGYTVALDMTARDFQDEAKKAGAPWFLAKSFDGSC 123

Query: 130 PVSAFVPKEKIPDPHKLKLWLKVNGLRQEG ETSSMIFSIPYIISYVSKIIITLEEGDIIL 189
P+ F+P IP+PH ++L+ K+NG+ +Q T MIF IP ++ Y ++ TLE GD++L
Sbjct: 124 PIGGFLPVSDIPNPHDVELFCKINGKDQQRCTDVMIFDIPTLLEYTTQFFTLEVGDVVL 183

Query: 190 TGTPKGVGPVKENDEIEAGIHGLVSMTEKVE 220
TGTP GV + D IE G+ ++ F V+
Sbjct: 184 TGTPAGVTKINSGDVIEFGLTDKLSKFNQV 214

Pedant information for DKFZphfk2_46j20, frame 1

Report for DKFZphfk2_46j20.1

[LENGTH] 224
[MW] 24843.07
[pI] 6.96
[HOMOL] PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans* 8e-55
[FUNCAT] r general function prediction [M. jannaschii, MJ1656] 9e-40
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YNL168c] 4e-38
[EC] 5.3.3.10 5-Carboxymethyl-2-hydroxymuconate delta-isomerase 1e-35
[PIRKW] isomerase 1e-35
[PIRKW] intramolecular oxidoreductase 1e-35
[SUPFAM] 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase 1e-46
[PROSITE] MYRISTYL 4
[PROSITE] AMIDATION 1

[PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 3
 [KW] Alpha_Beta

SEQ MGIMAASRPLSRFEWVGKNIVCVGRNYADHVREMRSAVLSEPVFLKPSTAYAPEGSPIL
 PRD cccccccccchhhhhccceeeecchhhhhhhhhccccceeecccccccccccccccc

SEQ MPAYTRNLHHELELGVVMGKRCRAVPEAAAMDYVGGYALCLDMTARDVQDECKKGLPWT
 PRD cccccchhhhhhheeeccccccccchhhhhhhheeeecchhhhhhhhhhhcccccc

SEQ LAKSFTASCPVSAFVPKEKIPDPHKLKLWLKVNGLRQEGETSSMIFSIPIYIISYVSKII
 PRD cccccccccceeeccccccccceeeccccccccccccceechhhhhhhhhhh

SEQ TLEEGDIILTGTGPKGVGPVKENDEIEAGIHGLVSMTFKVEKPEY
 PRD hccccceeeccccccccceeeeccccccccccccccc

Prosites for DKFZphfkd2_46j20.1

PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	192->195	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00008	2->8	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	116->122	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00009	78->82	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfkd2_46j20.1)

DKFZphfkd2_46k19

group: transcription factors

DKFZphfkd2_46k19.3 encodes a novel 130 amino acid protein similar to rat Dcoh, a bifunctional protein-binding transcriptional co-activator.

Dcoh is a bifunctional protein, complexed with biopterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the biopterin cofactor of phenylalanine hydroxylase.

The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

strong similarity to pterin-4-alpha-carbinolamine dehydratase

potential start at Bp 102 according to similar proteins,
both genomic sequences are from chromosome 5,

Sequenced by MediGenomix

Locus: map="5"

Insert length: 5641 bp

Poly A stretch at pos. 5617, polyadenylation signal at pos. 5598

```

1 CAGCCCTCGG CAGACGGCCA ATGGCGGCGG TGCTCGGGGC GCTCGGGGCG
51 AC CGCGGCGCT TGTTGGCGGC GCTGCGAGGC CAGAGCCTAG GGCTAGCGGC
101 CATGTCAATCA GGTACTCACA GGTTGATTGC AGAGGAGAGG AACCAAGCTA
151 TACTTGACCT TAAAGCAGCA GGATGGTCGG AATTAAAGTGA GAGAGATGCC
201 ATCTACAAAG AATTCTCCTT CCACAATTTT AATCAGGCAT TTGGCTTTAT
251 GTCCCGAGTT GCCCTACAAG CAGAGAAGAT GAATCATCAC CCAGAATGGT
301 TCAATGTATA CAACAAGGTC CAGATAACTC TCACCTCACA TGACTGTGGT
351 GAACTGACCA AAAAAGATGT GAAGCTGGCC AAGTTTATTG AAAAAGCAGC
401 TGCTTCTGTG TGATTCTTC CAAAATACAT AAGTCTGAGA GGCTAAACTT
451 GATGGCTGTG TTAACATATG TCACGTGTAG CACAGTGGAG AAAGCAGGAT
501 ATGGCTCATA ATGACAGTGG TGAAGACCTG CGAATGAACT TGCTAGTTAA
551 CACCTACATT AGGGTTTGAC ATAGGTCTAT GTTATGGGTC GCTGCATCTG
601 CTGGAATCA CAGACTTTAC TATAGAGAAT CAAAGATCCC GTATCCGAAG
651 TCTATGGAAA TGCTCATGGT GGTAATTTCC AACAGAATGA AACACCAAAC
701 TTGCTTAAAG TAACTCACGT TTCAATTTGA AAGAGATATT GTCAAAATTG
751 GAGGCCCCCA GGTTCCCTGTC TGTTCCAAAT CTTTGCATGA TGACAGTGGT
801 TTCTCTGATG TGGTAAGCTT TGGCTTTCTT CTGTTTTCTT TCTAAAAGAT
851 CACTGGAGTA GAGAGGAGTT AAACAGACAT GACCTTTGAC CTCTTGCATG
901 ACCTCCACAG ATAGCAAACC GGGCCGACAC ATGGTTGACG ATGTCCTTTT
951 CTACCAATGAA GTTAATGAAA GTTCTGAAAA TAGTGATTAC TTTCTGACAT
1001 TGATAGGATT TAGGAAACCT CTGGATAAAT AGCTTAAAGCA TGGCTGTTTA
1051 TGTTTTTGCT ATAGACAAAA AGCAGCAGCA TGTACATTGT ATTTGGACAC
1101 AAGCCTGCCT CGGTTAATAT ATTGAACAT TGGACCACTA GGGTTAGTAG
1151 GGAGCGGTCT GTACACTTTC TGATTACAGA TTCAGAAACA TTCTAGGTGG
1201 ACTCTGTAGC TTTCAGTTT GTAAAGTTAT CGGAAAAACA TCGGGAGGGT
1251 TTGGCCATCA TATGTGAGCT TTGTGTTTCA ATGCCAGTTA CTCAGGATTA
1301 GTAAATTAAT GACTGTCCAG AGGACTTCAG GGTCAACCAAG CTGCTGCACC
1351 TGCCATTGGC TGACTCTCCC CGGCTATCTG TGGCTGAGAT GGTGCTGCTT
1401 AGGTCACGCA GAGCATGAGC TGCTGCTGAA AGGGCACAGG AGATGGCCCT
1451 TGGGCTTCTC ATCCCAGGAT GCCTGCCCTG CCCACCAATC CATGAGAAGA
1501 TATGTATGAT TTCAGTAGGC CCTGGATCAG CTTGTCACTT CTGGTTTCCT
1551 GTTTGCTTTC CACTCACTCA GCTGGAGTTT CATTTCCAGA CTAAAGTCTT
1601 CATCATTTGC TTCAGAAACA GCATTATCT GTGGCTGTGC TGATGTAGTA
1651 CACCAAGAAG AACCTGGGCTC TTCTCTGTCA CTTTCAGTGG GCTACCTTCC
1701 CTCACCTCTC CAAGCAGCAT GAAAGAATTC TTTACATTTT TAATCTCTTT
1751 TTTGTTTTTC CCTGAAAGTA TGCTTTGGTG CTTAAAGAGA GAAGTCACAA
1801 AAGTATACTA CTGAGTTTCC TGGAGATGAA ATCCTGTTGT CCCTAGCTAT
1851 GTGAATGAGC ACAGGGATCC CTGATGCCAT TATTTTGAT ATTCATACGG
1901 CACACACTTA CTGAGGGCCT TCTGTGTGCC CTAGGGGATT GAGCACAGTG
1951 ACATATCAGG GCAGGTAGAA ACAGATGGAG AGCTGATGCG GGCTGTCTTA
2001 GAGCAGCTGC CCCAGGAGGC CCCTGTGGAT GGATGTTGGG CAGGAGCCCT
2051 GAGACGTTAG GGGCATATAA CTAAGGACA TAGCAGGAGT TATAGGAGGA
2101 GCTGATCCCT GAGGAAAACA ATGAAGACGG AGAAGATGGG GCTAAAGTTT
2151 GAATTGTGGG GACATTAATC ACGGTGATTC TTAAGACTTT GCTGTTGATG
2201 ATTTTAAATG GAGAAAATGA GTACGTAAGA TGTTATTTCC CAGTTCAGTA
2251 TATAGGTTGC CCACAAAGTA TTTTCTTACC ATGAATGGTC ATATATACTT
2301 GTGTGTAAT ACCAGGGACA GCAGAGATGG TGGGGTAGTT ACTTCTTTT
2351 CTTACAGCCC AAGAACTTTG GTGTCCAGGA GATTGACCAA TTTAGCCACT
2401 GAGCATTTAA TACAACACAG GGCTACCCAG ATCCCACCTGT CCTGATTTGC
2451 CCTGAAAGCC AAAGGAGTCA GGAGAAGGTG AGTGGGGTGA ATATATTAAT
2501 CCTGAGAGTT GAACAGAGCA AAAATCCCCTA TTACTTTTGT ACTTAAAAACA

```

```

2551 TCTCTGCCAC ATGTGCTCAC TCTTTATATT CTGTTTAGGT GGTTTATATG
2601 TGCACATCCC ATCCTATGCC TGCAGTTAGC CAACTCAGGG TTTATATTGC
2651 CTCCTTTCTT TTTTCTTTT TTTTTTTTTT TTTTAAAGAG TGGGGTCTCG
2701 TTCTGTGCATG CAGACTGGAG TGCAGTGGTG TGATCACAGC TCATTGTAAC
2751 CTCCAACGCC TGGACTGAAG TGATCCTCCT GCCTTGGCCT CTCTGGTAGC
2801 TGGGACTACA GGTGCATGCC ACCACACCCA CCTAATTTTT TTTATTTTAA
2851 TTTTGTGTAG AGACAGTCTC ACTATCTTGC TCGGGCTGGT CCTGAACCTC
2901 TGGGCTCAAG TTATCTTGCT GCCTCAGCCT CCCATGGGTA ATCTTTATTT
2951 CCTTTTTTTT TTTTTTTTGG AGATGGAGTT TCGCTCTTGT CGCCCAGGCT
3001 GGAGTGCAAT GGCACGATCT TGGCTCACTG CAGTCTCCAC CTCCTGGGTT
3051 CAGGTGATTC TCCATCCTCG GCCTACTGAG TAGCTGAGAT TACAGGCAAC
3101 TGCCACCATG CGCGGCTAAT TTGTGTATTT TTTTATAGTA AGAGATGGGG
3151 TTTGCGCATG TTGGCCGGAC TGGTCTTAGA CTCCTGACCT CAAGCGACCT
3201 GCCTGCCTTG GCCTCCCAA GTGCTGGGAT TACAGGCATG AGCCGCTATG
3251 CCTCGTCTGT GATTTTTATT TCTTATTTTT TTTTATAGAG TGGGGGTCTC
3301 ACTATGCTGC TCAGGCTGAT CTCAAACTCC TGGCCTCAAG TGATCCTCCC
3351 ACCTTAGCCT CCCAAGTTGC TGGGATTATA AGTGTGAGCC ACTATCCCTA
3401 CCTCACTATT ACCTTCTTTG CTTCCTTGTG TTTCTTTTGT TCTAAGTCAA
3451 ACCCATCACA ATCTTTTCTT GTCCTTCCAG GTGTTTTCCA GTGCTGTGCC
3501 CTGGATGTGC TCTCTTTCTC TTAGAGCCCA GAGAACTTGC TTTTCCCTCT
3551 TATATATGAC CCTTAACCTT TTCTAACACA TTATTAAGGG CCGTGTCTA
3601 TCAGCTGGGG GCACCTCTTG AAGGGAGGGC CTTTGTGTGG TCTGTTTCTA
3651 GTGACTTCCA GCTTTAACCC AGAGCCTCAT GATTGCTGGG TGCCCATAGC
3701 CTTTTTGTCT AATGGAGGCA CTCAGTCTCC TTGGGAAGAG AGAATCCATG
3751 ATAGACCCAC TTGGGAGCTC CCCACTTCAG GGGCCTACAC ACTGGTAATG
3801 CAACAGAAATG CCCAAGAGTG ACCTCATAAA GCAAGGATTC CCTTCGTGGC
3851 CCCTTCTCTG CTGCCTCTCA GAATCCAGAC GCTAAGGAAA ATCCCTAAGC
3901 AGAGATTTTC TGTGGATGC TAAAAGCAAG GAATAAAAGT TGAAAAATTTG
3951 GAAAATGTCT CAACACCGTC ACCAGCGCCA CTCGAGAGTC ATTTCTAGTT
4001 CACCAAGTGA CACTACATCG GTGGGATTTT GCCCAACATT CAAGAAATTT
4051 AAGTAAATAT TATCTATCTC CATTGCCTGT TAAGAAATGT GCTAGTAGAA
4101 GTGTGAGGGC AGGGTGTCTG TGTCTCTCTA GCCTCTTCCC TCAGATACTC
4151 GTCTGCTTAC CAAAATAAGT TGCATGTCCT TGACAATCTG GTTTCATGTA
4201 TTGGTGAGGC TGGCATGCTA TTACCTTTAT GTGCCCTGTA GACTTGAATG
4251 ACCAGTTTGA CCAGTTTGAC TGTAGATAAA TCAGAAAGCT TTTCTCTTTT
4301 TTTATAATAG ACCCATCTC AAATCAGATA ATGAAAATTA CATATCTTGA
4351 TATATTAGAA AAGTATATAC ATTCTGGCTG GGCACGGTGG CTCACGCCCTG
4401 TAATCCCTGC ACTTTGAGAG GCTGGGGCGG ATCACTTGAG GTCAGGAGTT
4451 TGAGACCGGC CTGGCCAGCG TGGCGAAACC CCATCTCTAC TAAAAATACA
4501 CAGATTAGCC CGGAGTGATG GTGTGCACCT GTTGTCCTAG CTAATCAGGA
4551 TGCTGAGGCA GGAGAATCCC TTTAACCTGG GGGGCGAAGG TTGCAGTGAG
4601 CCAGGATTGC ACCACTGCAC TCCAGCCTGG GTGACGGAAC GGGACTCTGT
4651 CTGAGAAAAA AAAAAAAGA AGAGGAAAAA GAAAAATATA TATTCATAT
4701 TTTTAAATCT TATGAGAAATG TGTTCATTTC ATTTGTAACA TATAATGGGA
4751 AACAGTAATA CGTACTCTGA GAAAAATTGC AAAGCACAGA TAAATGGAAA
4801 TAAACAGGAA AAAGAATCAC CTATAACCTC ACCATCCATA GACAGACACT
4851 GTTAAATTTT TGGCATATTT CCTGCTGATT TTTTCTACTG CTGATTTTTG
4901 CACAGGTGAG ATAATTTTGA ACAGAGAATT TTGTATCTTT GGTTTTTGTG
4951 TTTTCGCTGA CAAAAAACA AAAGATATAA AAATGGATCA TAAACATTTT
5001 TCTAAATCCT GAAAAGTGCA TAGACATATT TTAGTGCTCTG TATTTACAAA
5051 GATGGACATA CCATAATTTA CTTACACAGT CCTTTTTGTT AGATGTTTAA
5101 GTTGTTTTCA AGCTTCTCAG TGCTGGAAAA AATACTGAGA TAGACATGTT
5151 TAGTTGAAGT TATTTCAATT CAGGTTATAT TATCTTGGGT CAGAGAAATGA
5201 ATGGTTCTCA GGCTTTTCAA AAGAGCTGGT CAGTTTTTAT GCCTCTGGCA
5251 GTTTTTGAGA GTGCTCAATC ATACTACACT GTTGCCAGCA TTAGATCTTA
5301 TCACATTTAA GTCATTGCTA ATTTTATAAA CAAAAACAAT GGTTTTACTT
5351 TGCACTCTCC TGATTGGTGT TGCTGTAGAA CATATTTGGA GAAGTTTGTT
5401 TGTCTTTGGT GTTTATTCCA TGAATAGATT GTGTGCCCAT TTTCTCTTGG
5451 GGTATTACAG TTTTATTAC TGATGTGAGC ATCTGTATGG GTGATTATTT
5501 GATGATTATC AGTTTTGCTT AGTAGACTGG CAATATTTAG TCTTGTCTGC
5551 ACTGTGTCTC CAGTGCCAAC TAGATTGCTT GATATGTAGT TGCCACTCAA
5601 TAAAGATTTG TTGAGTCAAT GAAAAAATAA AAAAAAATAA A

```

BLAST Results

Entry AC004764 from database EMBL:
Homo sapiens chromosome 5, P1 clone 255g5 (LBNL H61), complete sequence.

Score = 11057, P = 0.0e+00, identities = 2217/2224
Bp 428-5625 of cDNA == Bp 2912-8107 of AC004764

Entry HSAC1555 from database EMBL:
Homo sapiens (subclone 1_d8 from BAC H75) DNA sequence, complete sequence.

Score = 575, P = 5.1e-30, identities = 115/115
Bp ~240- 430 of cDNA == HSAC1555 splice pattern

Medline entries

93186787:
 Phenylalanine hydroxylase-stimulating protein/pterin-4
 alpha-carbinolamine dehydratase from rat and human liver.
 Purification, characterization, and complete amino acid
 sequence.

93101632:
 Identity of 4a-carbinolamine dehydratase, a component of
 the phenylalanine hydroxylation system, and DCoH, a
 transregulator of homeodomain proteins.

95242099:
 Crystal structure of DCoH, a bifunctional, protein-binding
 transcriptional coactivator

Peptide information for frame 3

ORF from 21 bp to 410 bp; peptide length: 130
 Category: strong similarity to known protein

1 MAAVLGALGA TRLLAALRG QSLGLAAMSS GTHRLIAEER NQAILDLKAA
 51 GWSELSERDA IYKEFSFHNF NQAFGFMSRV ALQAEKMNHH PEWFNVYNKV
 101 QITLTSHDCG ELTKKDVKLA KFIEKAAASV

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_46k19, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfd2_46k19, frame 3

Report for DKFZphfd2_46k19.3

[LENGTH] 130
 [MW] 14377.56
 [pI] 9.17
 [HOMOL] PIR:A47189 pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) - rat 4e-34

[FUNCAT] 01.07.99 other vitamin, cofactor, and prosthetic group activities [S.
 cerevisiae, YHL018w] 5e-04
 [SCOP] dldchg_4.38.1.1.1 Pterin-4a-carbinolamine dehydratas 4e-50
 [EC] 4.2.1.96 Tetrahydrobiopterin dehydratase 6e-34
 [PIRKW] nucleus 6e-34
 [PIRKW] carbon-oxygen lyase 6e-34
 [PIRKW] homotetramer 6e-34
 [PIRKW] hydro-lyase 6e-34
 [PIRKW] cytosol 6e-34
 [PIRKW] acetylated amino end 6e-34
 [PIRKW] homodimer 6e-34
 [SUPFAM] pterin-4-alpha-carbinolamine dehydratase 6e-34
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 3
 [PROSITE] PKC_PHOSPHO_SITE 4
 [KW] Alpha_Beta
 [KW] 3D
 [KW] LOW_COMPLEXITY 14.62 %

SEQ MAAVLGALGATRLLAALRGQSLGLAAMSSGTHRLIAEERNQAILDLKAAGWSELSERDA
 SEG .xx
 ldchBCCCCCHHHHHHHHHHHHHHHHHHCCEEECCCE

SEQ IYKEFSFHNFNQAFGFMSRVALQAEKMNHHPEWFNVYNKVQITLTSHDCGELTKKDVKLA
 SEG
 ldchB EEEEECCCHHHHHHHHHHHHHHHHHHHCCCCCEEEETTTEEEECBTTTBTCCCHHHHHH

SEQ KFIEKAAASV

SEG
ldchB HHHHHHHHHH

Prosites for DKFZphfkd2_46k19.3

PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	32->35	PKC_PHOSPHO_SITE	PDOC00005
PS00005	56->59	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	113->117	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2_46k19.3)

DKFZphfkd2_46m4

group: signal transduction

DKFZphfkd2_46m4.3 encodes a novel 198 amino acid putative GTP-binding protein related to the SAR-1 family of Ras superfamily members.

SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

The new protein can find clinical application in modulating the transport of vesicles to the Golgi Apparatus, thus enabling post-translational modifications of the vesicles contents. Blocking of the molecule is expected to result modulation/blocking of secretory pathways.

nearly identical to mouse GTP-binding protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="438.9 cR from top of Chr10 linkage group"

Insert length: 2996 bp

Poly A stretch at pos. 2969, polyadenylation signal at pos. 2958

```

1 ACATCCGGCG AGTAGCTGGC GGTCCCGGGT GCTGCTGGTT AGTGTGCTCT
51 GAGGGAGGGT CCGAGCCAGC CGCTGTTTTG CCGGAGGAGC CCCTCAGGCC
101 GTAGTAAGCA TTAATAATGT CTTTCATCTT TGAGTGGATC TACAATGGCT
151 TCAGCAGTGT GCTCCAGTTC CTAGGACTGT ACAAGAAATC TGGAAAACCT
201 GTATTCTTAG GTTTGGATAA TGCAGGCAAA ACCACTCTTC TTCACATGCT
251 CAAAGATGAC AGATTGGGCC AACATGTTC AACACTACAT CCGACATCAG
301 AAGAGCTAAC AATTGCTGGA ATGACCTTTA CAACCTTTGA TCTTGGTGGG
351 CACGAGCAAG CACGTCGCGT TTGGA AAAAT TATCTCCAG CAATTAATGG
401 GATTGCTCTT CTGGTGGACT GTGCAGATCA TTCTCGCCTC GTGGAATCCA
451 AAGTTGAGCT TAATGCTTTA ATGACTGATG AAACAATATC CAATGTGCCA
501 ATCCTTATCT TGGGTAACAA AATTGACAGA ACAGATGCAA TCAGTGAAGA
551 AAAACTCCGT GAGATATTG GGCTTTATGG ACAGACCACA GGAAAGGGGA
601 ATGTGACCCG GAAGGAGCTG AATGCTCGCC CCATGGAAGT GTTCATGTGC
651 AGTGTGCTCA AGAGGCAAGG TTACGGCGAG GGTTCGCGCT GGCTCTCCCA
701 GTATATTGAC TGATGTTTGG ACGGTGAAAA TAAAAGAGTT TACTTCTCT
751 GGACTGATCC TATTCACAGC TTCTCATGA ACTTTCTTAA TAGAACAAGG
801 ATAGCTCTCC AACCATGTCT GGCCTTGAGA AGCCAAGAGT CTCTGTCAAC
851 TCTCTCATTT CCCAGTGGTG ACATGTGCTC TTCTCCACAC TGTGGGAGG
901 TAATGCTGCC CCACGTGCTG GTGCAGGTCA GTATCCTGGG ACTTGGGAAGC
951 TGGCAGGATT TGCCGGGTAA AGCTGTATGC CATCATGGGG CACCTGAAAA
1001 GAAAAACACG TCTCACCCT GTGGTTGATT CAAAAGAAAG TGATTCTATT
1051 TTTTAAAGAA AGCGTTGTTA ATGTAATTGG TATCCCTCCT AACTTTTTGA
1101 GTTCACAATT TACTGGTCC AGAGTTTTCT ATTCTTTTTT TTTTTTAA
1151 CTAATGAATG ACATTAGAT ACTTCATAAA ATTATGAACA GATATGGAGG
1201 CCAGAGCTCA TTTGGGTAAA CTTACTCTCT CTGAGTTAGC AGGTTGGTGA
1251 GAGAACTCCG CCTGAGCTCA CCGTCTCTC TGACTGCCTT GGAGTAGGTG
1301 GCATAACCTT GTGCACAGAG AACTAGAAAA GGGGCAGAAC CCCGGCCTTG
1351 CAGTTGTGGC AGGTTTCCAC TGTGGTAAGC TAGGTTCAAT CCTCATCAAG
1401 GAATGTGTAG CAGATTGTTT ACTGTGGAGG AGGTAATTAT AGAATGGGTT
1451 ATTGTTGTTA TTCTTACTCA TGAAGTTACA GATTTTAGCC AGTCTTTGCT
1501 TTTTACTTTT TGTGAAATTT AATTCTCTC TATAGCCTT TCCTTTTTCG
1551 TTTTCAGTTA TCAAAAGTGA CTTTGACCTC ATAAGAGAGT TGAGAACATC
1601 TCTCGTGCTA CATACTGCAG GTGCATCAGT TACTTTTGCA CAGATTCTAG
1651 GGGGACATTT TTCTGAATAG GAAGACAGGA CAAAGTTAAC AGCTTAAGGG
1701 CTCTTAATTC TGTGAGTTGA GGACTTAAAA GTATTGTAGC ATTTGTTTGG
1751 ATCCATGAAA AATGTATTCA GTGGGCTTTA AAATTTCCAT TTGCAGAATT
1801 TGGTCTCTCA GGCTGTTTGG GAGCTCTTTT TTTTACATTT TTTCTCCTTT
1851 GACACCTATT TTATTGGTGT TTAAGTAAA GGTTAACATC TGAGCTTTT
1901 CCAGGTTTTT TTTTTTTTTT TTGATATGAA ATTGCTTTT TCCATTGCAG
1951 AAATAAGCTA GGGAAACACT AACCCAAAAA CTTTCTGTAG AGCTGTTCTT
2001 TTGGAGGCAG CATCACTTAT TGGCAGTAAA GACTCAGTAT AAAAGCACCA
2051 GCATCCCTAC TTGGGTGATG GGGATTAATT TTATAGCATT CCATTTTCCT
2101 AGTGCCACAT GTGAAATTGG ATTTTGATGA TCTTAATCTA TATTCTACCC
2151 TTATAATAAA AGATCAAAAG ATATATCTCC TATGAACAGA TTGGAGATAG
2201 GAGATGAAAA GTTGGGAGGA TGCCTTTATT CTAATGTGAG GGTAGGGAAA
2251 ATGTGGATTA CATTACTGGG GTGAAGGAGG CATTGTTCTT TAGTTGGAGT
2301 TCTCATTTTT ATTCTCCAGT ACTGACTTGT GGGGAAAGCA TACTTTTTCA
2351 CTGCCAGGTA CTGAATGCAG AGGCTCAGTG AAGTATATAT GTGGGAAGTG
2401 CATGCATTTC GTTTATTAGC AAACATAGCT GGATTAAGAC GAAGTTGTTG
2451 GTTTGGAAAG GGGTTAAAGC CTTAAGTGAA CAAATCTAGC TAACAGTGAA
2501 TGAAGTAGGT AATATAACTT GCATATTTTT AATTCCTTT GGTAAAGGT
2551 CCCCCATACT TCTCTGTTTG GAGACATGAG AAGTATGATT ACTTCAGTGT

```

```

2601 TAGTTTCTT AATTTTTTTT TTCCCCTATT TGTCCCTTGT CACTTTGTTG
2651 CAAGCTAGAA ATCTGTGGGT TATACATAGG GCAGCTCTTT GCGAAAGTGG
2701 TTTATTCCAC TGGAGAAAGG GGATTGAAA TCAGTTAGAA CCAATGTATT
2751 TC TTGCCCCA CGGAACACTA TTCCTATAAG ATAGCTGAAA GAAGCTGCTG
2801 TGAGGAGCTC AGCTCCAACA CAGGATCAGC ACCTTGATA GGAATTCCCA
2851 TGAATTATGA CTTCTCATTG TGTTTTATCA GAGTGCATAT ATGTCCTACT
2901 TCAGGAAAAG TAAAACAGTC ATTTACGAAA GAAAGTCAAT CTGTATCCTA
2951 AGCATTTTAA TAAAAGTTA AAACAAAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

```

-----
Entry HS679348 from database EMBL:
human STS WI-16722.
Length = 265
Minus Strand HSPs:
Score = 1242 (186.4 bits), Expect = 2.8e-50, P = 2.8e-50
Identities = 260/265 (98%)

```

Medline entries

```

-----
94085558:
Molecular analysis of SAR1-related cDNAs from a mouse
pituitary cell line.

```

Peptide information for frame 3

```

-----
ORF from 117 bp to 710 bp; peptide length: 198
Category: strong similarity to known protein

```

```

1 MSFIFEWIYN GFSSVLQFLG LYKKSGLVF LGLDNAGKTT LLHMLKDDRL
51 GQHVPTLHPT SEELTIAGMT FTFDLGGHE QARRVWKNYL PAINGIVFLV
101 DCADHSRLVE SKVELNALMT DETISNPIL ILGNKIDRTD AISEEKLREI
151 FGLYGQTTGK GNVTLKELNA RPMEVFMCSV LKRQGYGEGF RWLSQYID

```

BLASTP hits

```

Entry S39543 from database PIR:
GTP-binding protein - mouse
Length = 198
Score = 1029 (362.2 bits), Expect = 5.1e-104, P = 5.1e-104
Identities = 197/198 (99%), Positives = 198/198 (100%)

```

```

Entry SARA_MOUSE from database SWISSPROT:
GTP-BINDING PROTEIN SARA.
Length = 198
Score = 1012 (356.2 bits), Expect = 3.2e-102, P = 3.2e-102
Identities = 195/198 (98%), Positives = 196/198 (98%)

```

```

Entry CEZK180_4 from database TREMBL:
gene: "ZK180.4"; Caenorhabditis elegans cosmid ZK180.
Length = 193
Score = 679 (239.0 bits), Expect = 6.3e-67, P = 6.3e-67
Identities = 125/197 (63%), Positives = 161/197 (81%)

```

Alert BLASTP hits for DKFZphfkd2_46m4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_46m4, frame 3

Report for DKFZphfkd2_46m4.3

```

[LENGTH]      198
[MW]           22367.00
[pI]           6.21
[HOMOL]       PIR:S39543 GTP-binding protein - mouse 1e-112

```

```

[FUNCAT]      08.07 vesicular transport (golgi network, etc.)      [S. cerevisiae, YPL218w]
1e-58
[FUNCAT]      30.09 organization of intracellular transport vesicles      [S. cerevisiae,
YPL218w] 1e-58
[FUNCAT]      06.10 assembly of protein complexes      [S. cerevisiae, YOR094w] 2e-23
[FUNCAT]      06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing)      [S. cerevisiae, YPL051w] 4e-22
[FUNCAT]      30.08 organization of golgi      [S. cerevisiae, YDL192w] 3e-20
[FUNCAT]      30.03 organization of cytoplasm      [S. cerevisiae, YBR164c] 3e-19
[FUNCAT]      03.22 cell cycle control and mitosis      [S. cerevisiae, YMR138w] 2e-09
[FUNCAT]      30.04 organization of cytoskeleton      [S. cerevisiae, YMR138w] 2e-09
[FUNCAT]      98 classification not yet clear-cut      [S. cerevisiae, YHR168w] 7e-05
[FUNCAT]      30.02 organization of plasma membrane      [S. cerevisiae, YHR005c] 1e-04
[FUNCAT]      30.07 organization of endoplasmatic reticulum      [S. cerevisiae, YKL154w]
1e-04
[FUNCAT]      03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHR005c] 1e-04
[FUNCAT]      10.05.07 g-proteins      [S. cerevisiae, YHR005c] 1e-04
[FUNCAT]      06.04 protein targeting, sorting and translocation      [S. cerevisiae, YKL154w]
1e-04
[FUNCAT]      08.19 cellular import [S. cerevisiae, YML001w] 3e-04
[BLOCKS]      BL00395A Alanine racemase pyridoxal-phosphate attachment site proteins
[BLOCKS]      BL01019B ADP-ribosylation factors family proteins
[BLOCKS]      BL01019A ADP-ribosylation factors family proteins
[BLOCKS]      BL01020D SAR1 family proteins
[BLOCKS]      BL01020C SAR1 family proteins
[BLOCKS]      BL01020B SAR1 family proteins
[BLOCKS]      BL01020A SAR1 family proteins
[SCOP]        d1plj_ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens)] 7e-36
[SCOP]        d1guaa_ 3.25.1.3.10 Rap1A [Human (Homo sapiens)] 8e-40
[SCOP]        d1rrf_ 3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) [rat (Rattus)] 2e-55
[SCOP]        d1hurb_ 3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) [human (Homo sapiens)] 1e-58
[SCOP]        d1gota2_ 3.25.1.3.3 (1-54,171-326) Transducin (alpha subunit) [Rattus] 2e-33
[SCOP]        d1tadb2_ 3.25.1.3.2 (1-30,152-316) Transducin (alpha subunit) 6e-36
[PIRKW]       glycoprotein 4e-19
[PIRKW]       monomer 1e-16
[PIRKW]       P-loop 3e-64
[PIRKW]       lipoprotein 4e-19
[PIRKW]       GTP binding 3e-64
[SUPFAM]      ADP-ribosylation factor 5e-22
[PROSITE]     ATP_GTP_A 1
[PROSITE]     MYRISTYL 3
[PROSITE]     SAR1 1
[PROSITE]     CK2_PHOSPHO_SITE 4
[PROSITE]     PKC_PHOSPHO_SITE 3
[PROSITE]     ASN_GLYCOSYLATION 1
[PFAM]        ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
[KW]          Alpha_Beta
[KW]          3D

```

```

SEQ      MSFIFEWIYNGFSSVLQFLGLYKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPTLHPT
lhurA    .....TTTTTCCCEEEEEETTTTCHHHHHHHHCCCEEEEEETTEE

```

```

SEQ      SEELTIAGMTFTTDFDLGGHEQARRVWKNYLPAINGIVFLVDCADHSRLVESKVELNALMT
lhurA    EEEEEETEEEEETTTTTTCCCHHHHHHCEEEEEETTTTHHHHHHHHHHHHHHH

```

```

SEQ      DETISNVPIILGNKIDRTDAISEEKLREIFGLYGQTTGKGNVTLKELNARPMVEVFMCSV
lhurA    TTTTTTTEEEEEETTTTTTCCCHHHHHHHHCGG.....

```

```

SEQ      LKRQGYGEGFRWLSQYID
lhurA    .....

```

Prosites for DKFZphkd2_46m4.3

PS00001	162->166	ASN_GLYCOSYLATION	PDOC00001
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	60->64	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	111->115	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	68->74	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00017	32->40	ATP_GTP_A	PDOC00017
PS01020	171->197	SAR1	PDOC00782

Pfam for DKFZphkd2_46m4.3

HMM_NAME	ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)		
HMM	*GMgWfsIFrkmWGLWNKEMRILMLGLDNAGKTTILYMLKlgEIVTTIPT		
	++ FS+++++GL++K+++++LGLDNAGKTT+L+MLK++++ ++PT		
Query	9	-YNGFSSVLQFLGLYKKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPT	56
HMM	IGFNVETVeYKNIKFNVWDVGGQdsIRPYWRHYYPNTDGIWVVDsADRD		
	++++E++++ ++F+++D+GG++++R++W++Y P+++GI+++VD+AD++		
Query	57	LHPTSEELTIAGMTFTTFFDLGGHEQARRVWKNYLPAINGIVFLVDCADHS	106
HMM	RMeEaKqELHaMLNEEELrDAPLLIFANKQDLPGAMsEsEIREaLGLHeI		
	R+ E+K+EL+A++++E ++++P+LI++NK+D+ +A+SE+++RE+ GL+ +		
Query	107	RLVESKVELNALMTDETISNVPIILGNKIDRTDAISEEKLREIFGLYGQ	156
HMM	RCn.....RPWYIQMCCAvtGEGLYEGMDWLSNYInkRkK*		
	+++ RP++++MC+++++G++EG++WLS+YI		
Query	157	TTGKGNVTLKELNARPMEVFMCSVLKRQGYGEGFRWLSQYI-----	197

DKFZphfkd2_47a4

group: transcription factor

DKFZphfkd2_47a4.1 encodes a novel 280 amino acid protein with similarity to zinc finger proteins.

The new protein is a putative transcription factor with one C2H2 zinc fingers.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to C.elegans F46B6.7

potential frame shift at 1092, will be checked see BLASTX

Sequenced by MediGenomix

Locus: map="7q31"

Insert length: 1756 bp

Poly A stretch at pos. 1737, no polyadenylation signal found

```

1 CCCTTTTCTT TTCTGCCGGG TAATGGCTGC TTCCAAGACC CAGGGGGCTG
51 TCGCCCCGAAT GCAGGAAGAC CGTGATGGGA GCTGCAGCAC AGTCGGGGGT
101 GTAGGTTATG GGGTAAGGAT TGTATCCTGG AGCCGCTTTC CCTGCCAGAA
151 AGTCCAGGTG GCACCACCAC TTTAGAAGGT TCTCCATCTG TGCCTTGTAT
201 TTTCTGTGAA GAACATTTTC CTGTGGCTGA ACAAGACAAA CTTCGGAAGC
251 ACATGATTAT TGAGCATAAG ATTGTCATAG CTGATGTCAA GTTGGTTGCT
301 GATTTCCTAAA GGTACATTTT ATATTGGAGG AAAAGGTTCA CTGAACAGCC
351 CATCACAGAT TTTTGTAGTG TAATAAGAAT TAATTCCACT GCTCCATTTG
401 AAGAACACAGA GAATTATTTT TTGTTATGTG ACGTTTACC AGAAGATAGA
451 ATTCTTAGAG AAGAGCTTCA GAAACAGAGA CTGAGAGAAA TTCTGGAACA
501 ACAGCAGCAA GAACGAAATG ATAACAATTT TCATGGCGTT TGTATGTTTT
551 GCATGAAGA ATTCTTGGA AACAGATCTG TTATTTTGAA CCACATGGCC
601 AGAGAACATG CTTTCAACAT TGGATTGCCA GACAACATTG TAAACTGCAA
651 TGAATTTTTG TGTACATTAC AGAAAAAGCT TGACAATTTG CAGTGTCTGT
701 ACTGTGAGAA GACCTTCAGG GGCAAAAAATA CACTTAAAGA TCACATGAGG
751 AAAAAACAGC ATCGTAAGAT TAATCCTAAG AACAGAGAAAT ATGACAGATT
801 TTATGTCATC AATTATTTGG AACTTGGAAG ATCGTGGGAG GAAGTTCAGT
851 TGGAAGATGA TCGGGAGTTG CTGGACCATC AGGAAGATGA CTGGTCTGAT
901 TGGGAAGAAC ACCCTGCCTC TGCAGTCTGC TTATTTTGTG AAAAGCAAGC
951 AGAAACAATT GAGAAAGTTG ATGTCCACAT GGAGGATGCA CACGAATTTG
1001 ATCTTCTCAA AATAAAGTCA GAACCTGGAT TAAATTTCTA TCAGCAAGTG
1051 AAACCTGGTCA ATTTTATTCG GAGGCAAGTT CACCAATGCA GATGATGGCT
1101 GCCATGTGAA GTTCAAATCC AAAGCAGACT TAAGAACTCA CATGGAAGAA
1151 ACTAACAACA CTTGCTGCTC CCCCAGTAGA AAGACGTGGG ATCAACTGGA
1201 GTATTATTTT CCAACCTATG AAAATGACAC TCTCCTGTGT AACTATCTG
1251 ACAGTGAAAG TGACCTGACA GCTCAGGAAC AAAATGAAAA TGTTCCTATC
1301 ATCAGTGAAG ATACATCTAA ACTGTATGCT TTGAAACAAA GCAGTATTTT
1351 GAACCAAGTT CTAATAAAG AGTACTTGAA AACCTAGAAG AAACCTACCAC
1401 AGAAGCAATT TTTTATGTTT TTCTCCTATG AGACAGATAT GAAAGAACAA
1451 TTTAAATTTG AACATCAACA AAAGATTGGT CCTTGGTGAA ATAACTTTT
1501 CAAAAATGAA TGTCTTTTTC AAAAAATAAA GTAGAAAAAT GCACTTACTA
1551 AGAACATGAA AAAAAATGA AGTAGGAAAA TAAGATGAAG ACTTTGTATT
1601 TTGGCTGTAA AGTTTATTTG TGTGATCATC TTAAATTATC TCACCTCAT
1651 AAACCTATAA TTATATATAG AAGTATATGT CAATTACAAA GAAATGAAAT
1701 GTTCAAATTA TTTATAAACC TGATTTTTC AATCAGCGAAA AAAAAA
1751 AAAAAA

```

BLAST Results

Entry AC004112 from database EMBL:

Homo sapiens BAC clone RG313E03 from 7q31, complete sequence.

Score = 2660, P = 3.0e-241, identities = 534/535

> 10 exons

Entry AC004111 from database EMBL:

Homo sapiens BAC clone RG103H13 from 7q31, complete sequence.

Score = 598, P = 5.8e-17, identities = 128/137

1 exon

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 253 bp to 1092 bp; peptide length: 280
Category: similarity to unknown protein

```

1 MIEHKIVIA DVKLVDQFQ YILYWRKRF EQPITDFCSV IRINSTAPFE
51 EQENYFLLCD VLPEDRILRE ELQKQRLREI LEQQQQERND NNFHGVCMFC
101 NEEFLGNRSV ILNHMAREHA FNIGLPDNIV NCNEFLCTLQ KKLDNLQCLY
151 CEKTFRGKNT LKDHMRKKQH RKINPKNREY DRFYVINYLE LGKSWEEVQL
201 EDDRELLDHQ EDDWSDWEEH PASAVCLFCE KQAEETIEKLY VHMEADAHEFD
251 LLKIKSELGL NFYQQVKLVN FIRRVQHOCR

```

BLASTP hits

Entry CEF46B6_6 from database TREMBLNEW:
product: "F46B6.7"; Caenorhabditis elegans cosmid F46B6
>TREMBL:CEF46B6_6 product: "F46B6.7"; Caenorhabditis elegans cosmid F46B6
Score = 630, P = 1.1e-61, identities = 123/289, positives = 183/289

Entry AF059531_1 from database TREMBLNEW:
gene: "PRMT3"; product: "protein arginine N-methyltransferase 3"; Homo sapiens protein arginine N-methyltransferase 3 (PRMT3) mRNA, partial cds. >TREMBL:AF059531_1 gene: "PRMT3"; product: "protein arginine N-methyltransferase 3"; Homo sapiens protein arginine N-methyltransferase 3 (PRMT3) mRNA, partial cds.
Score = 120, P = 1.5e-04, identities = 23/78, positives = 42/78

Entry YB9M_YEAST from database SWISSPROT:
34.7 KD PROTEIN IN SHM1-MRPL37 INTERGENIC REGION.
Score = 112, P = 4.6e-04, identities = 43/165, positives = 71/165

Alert BLASTP hits for DKFZphfkd2_47a4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_47a4, frame 1

Report for DKFZphfkd2_47a4.1

```

[LENGTH]      280
[MW]           33921.94
[PI]           5.63
[HOMOL]        TREMBL:CEF46B6_5 gene: "F46B6.7"; Caenorhabditis elegans cosmid F46B6 1e-56

```

```

[BLOCKS]       BL01032B Protein phosphatase 2C proteins
[BLOCKS]       BL00028 Zinc finger, C2H2 type, domain proteins
[PROSITE]      MYRISTYL 1
[PROSITE]      ZINC_FINGER_C2H2 1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 3
[PROSITE]      TYR_PHOSPHO_SITE 2
[PROSITE]      PKC_PHOSPHO_SITE 2
[PROSITE]      ASN_GLYCOSYLATION 2
[PFAM]         Zinc_finger, C2H2 type
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY 8.21 %

```

```

SEQ  MIEHKIVIA DVKLVDQFQYILYWRKRFTEQPITDFCSVIRINSTAPFE EQENYFLLCD
SEQ  .....
PRD  cccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeeccccccchhhhhheeeccc

SEQ  VLPEDRILREELQKQRLREILEQQQQERNDNNFNHGVCMFCNEEFLGNRSVILNHMAREHA
SEQ  .....
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeeccccccceehhhhhhhhh

SEQ  FNIGLPDNIVNCNEFLCTLQKKLDNLQCLYCEKTFRGKNTLKDHMRKKQHRKINPKNREY

```

SEG
PRD hccccccccchhhhhhhhhhhhhheeeccccccchhhhhhhhhcccccccc

SEQ DRFYVINYLELGKSWEEVQLEDDRELLDHQEDDSDWEEHPASAVCLFCEKQAETIEKLY
SEG
PRD ceeeeeeeeccccchhhhhhhcchhhhhccccccccccccccccchhhhhhhhhhhhh

SEQ VHMEDAHEFDLLKIKSELGLNFYQQVKLVNFIRQVHQCR
SEG
PRD hhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhcccc

Prosite for DKFZphfd2_47a4.1

PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	107->111	ASN_GLYCOSYLATION	PDOC00001
PS00004	27->31	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	160->164	CK2_PHOSPHO_SITE	PDOC00006
PS00006	194->198	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00007	178->185	TYR_PHOSPHO_SITE	PDOC00007
PS00007	13->22	TYR_PHOSPHO_SITE	PDOC00007
PS00008	124->130	MYRISTYL	PDOC00008
PS00028	148->171	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKFZphfd2_47a4.1

HMM_NAME	Zinc finger, C2H2 type		
HMM	*CpwPDCgKtFrrwsNLrRHMR..T.H*		
	C + C+KTFR + +L+ HMR H		
Query	148	CLY--CEKTFRGKNTLKDHRKK-QH	170

DKFZphfkd2_4b6

group: kidney derived

DKFZphfkd2_4b6 encodes a novel 133 amino acid protein with similarity to Homo sapiens clone 25003 partial CDS.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to Homo sapiens clone 25003

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1936 bp

Poly A stretch at pos. 1916, polyadenylation signal at pos. 1890

```
1 GGGAGACTTG CAATGAAGTT AGAATGAACA GGAGGAGTCT GCAGCTTTTC
51 AGTGCCTGGG ATAACATATAG TTAAAGATC ATTGTGTAAA ATAGGATTTT
101 TAGTCAGCAT GCATTGTTTT AAACCGACTA ACTGATAGCC TAAACTTTTA
151 TTTTTCGATT TTGCCAATCC TTGGAGTTT GTTTTCGAGA ATTAAGAAAA
201 AAATGAATGT ATGATCATCT GAAAAGGGCT TTCTCTCAAT CCCACTTCAT
251 GGCATGACCT CTGCTGGATC ATTAGTTCTA GCCAGAGAAG TAGCAAAGGA
301 ACATGACGTC TGAGACCTCC CTCCCTCAT CAGTGGGGCT GACTGAGCTG
351 GGGGCTTGAA GCCCGAGGTA ACCTTTCCTG TCGAATGTTT CTTTAGAGAA
401 TGGCAATGGT CTCTGCGATG TCCTGGGTCC TGTATTGTG GATAAGTGCT
451 TGTGCAATGC TACTCTGCCA TGGATCCCTT CAGCACACTT TCCAGCAGCA
501 TCACCTGCAAC AGACCAGAAG GAGGGACGTG TGAAGTGATA GCAGCACACC
551 GATGTTGCAA CAAGAATCGC ATTGAGGAGC GGTCACAAAC AGTAAAGTGT
601 TCCTGTCTAC CTGGAAAAAGT GGCTGGAACA ACAAGAAACC GGCTTCTTTG
651 CGTCGATGCC TCCATAGTGA TTGGAAATG GTGGTGTGAG ATGGAGCCTT
701 GCCTAGAAGG AGAAGAATGT AAGACACTCC CTGACAATTC TGGATGGATG
751 TGCGCAACAG GCAACAAAAT TAAGACCACG AGAATTCACC CAAGAACCTA
801 ACAGAAGCAT TTGTGGTAGT AAAGGAAAC CAACCTCTG GAAAATACAT
851 TTTGAGAATC TCAAACATCT CACATATATA CAAGCCAAAT GGATTTCTTA
901 CTTGCACCTT GACTGGCTAC CAGATAATCA CAGTGCCTTT AGTGTGTGTA
951 ACGAAATATC CTACAGTGAG AAGACACAGC GTTTTGGCAT CACCATGGAA
1001 AGTGGGCTTA AAAAAGGGTC TTCTCAGTGA AATTTTGGG CATCATGAAG
1051 AACGATCAAC TATCTTCTAA TTTGAATCTA TAGTTACTTT GTACCATTG
1101 AAATATATGT ATATATATAT ATATAATATT TTGAAATATT ATCTATTCTC
1151 TTCAAGAAAT GAACAGTACC ACAGTTTGTG ACGGCTGGTG TACCCCTTTG
1201 AGTTTTGGAT GTTTTGTCTG TTTTGTCTTG TTTTGTAGT CATTCTTTT
1251 TCTAACGGCA AGGAAGATAT GTGCCCTTTT GAGAATTCAA GATGGCACTG
1301 ACACGGGAAG GCCAGCTACA GGTGGACTCC TGGAATTTGA GGCATCATAA
1351 TGATACTGAA TCAAGAACTT CCTTCTGCTT CTACCAGATG GCCCAAGGAA
1401 GCACATCGTC CTGTTTTATT GCTTTCTACC CTGTGCAATA TTAGCATGCA
1451 AGCTTGGCTT ACATAGTCAT ACTTTATATT CAATTGATAT ATAATAACCG
1501 TTCTAACCTC TTCCAGGAAA ATATTTTGTG AACTACTAGC TTTTCCACTT
1551 AGAAGAAAAT GAGGATTCCT AAGGGAGCCA CTCCACCATG CTATTAAGAC
1601 TCTGGCAGAG TTATGGGTAG GATATGGATC CCTACATGAA TAAGTCCCTG
1651 AAATACAATG TCTTAAGGCT TTGTATAGCT GTCTTAGACT GCAGAAATGT
1701 CCTCTGATTA AATCCAAAGT CTGGCATCGT TAACTACATA GTGCTGTAGC
1751 AACAAGTCTT ATCATGGCAT CTCTTCTAT GTTTGGTTTG CTTTTTCCAA
1801 GAGTATTCAG GTCTCCTCTT GTGAGATAGG AAGGCCATGA AAACAATTAG
1851 ATTTCAAGAT GATCTATGTG ACCAAATGTT GGACAGCCCT ATTAAAGTGG
1901 TAAACAACCT CTTTCTAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

References are listed in the following table:

1 MAMVSAMSWV LYLWISACAM LLCHGSLQHT FQQHHLHRPE GGTCEVIAAH
51 RCCNKNRIEE RSQTVKCSCL PGKVAGTTRN RPSCVDASIV IWKWWCMEPE
101 CLEGECKTL PDNSGWMCAT GNKIKTTRI H PRT

No BLASTP hits available

```
>TREMBLNEW:AF131851_1 product: "Unknown"; Homo sapiens clone 25003 mRNA
sequence, partial cds.
Length = 165
```

Query: 102 LEGEECTLPDNSGWMCAT-GNKIKTTTRI 129
LEGE C L + SGW C G +IKTT +
Sbjct: 136 LEGEGCDLLINRSWGTCTOPGGRIKTTTV 164

```
[LENGTH]      133
[MW]           15030.64
[PI]           8.49
[HOMOL]        TREMBLNEW:AFI31851_1 product: "Unknown"; Homo sapiens clone 25003 mRNA
sequence, partial cds. 4e-20
[KW]           Alpha Beta
[KW]           SIGNAL PEPTIDE 26
```

```

SEQ      MAMVSAMSWLWLYLWISACAMLLCHGSLQHTFQQHHLHRPEGGTCEVIAAHRCCKNKRIEE
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhccccccccceeeeeccccchhhh

SEQ      RSQTVKCSCLPGKVAGTTRNRPSCVDAISIVKWWCEMEPCLEGECKTLPDNSGWMCAT
PRD      hhhhhhhccccccccccccccccccccceeeeehhhhhhhhccccccccceeeeeccccceeeec

SEQ      GNKIKTTRIHPRT
PRD      cccccccccccccc

```

(No Pfam data available for DKFZphfkd2 4b6.1)

DKFZphfkd2_4c8

group: kidney derived

DKFZphfkd2_4c8 encodes a novel 153 amino acid protein with partial similarity to huntington's associated protein HAP1.

The novel protein contains a leucine zipper involved in protein-protein interaction.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KIAA0549 and HAP1

potential frame shift at Bp -1350-1500 will be checked

Sequenced by GBF

Locus: unknown

Insert length: 3182 bp

Poly A stretch at pos. 3162, polyadenylation signal at pos. 3135

```
1 GGGCTTCCCC CATAGAATTT TTCTTTTCAT TGCCCACTTT ACTGTTTTGG
51 CTCCAGACTG TCGTTAAGAA TGTACAGCCT AATTCTGGTG TGTTCGGGA
101 TATTCTTCTG TCCAGTATTC TGGAAAGGCG GGGAGGCATG GCAGCGTTTT
151 ACTTGACGTT GATGGTGTG TGAAGTCCAT TCTTTCCTCT GCAAGACTAC
201 TGACTATGCA GAAATTTATC GAAGCGGATT ATTATGAAC AGACTGGTAT
251 TATGAAGAAT GCTCGGATGT TTTATGTGCT GAAAGAGTTG GCCAGATGAC
301 TAAGACATAT AATGACATAG ATGCTGTCTC TCGGCTTCTT GAGGAGAAAG
351 AGCGGGGATT AGAATTGGCC GCTCGCATCG GCCAGTCGTT GTTGAAGAAG
401 AACAAAGACCC TAACCGAGAG GAACGAGCTG CTGGAGGAGC AGGTGGAACA
451 CATCAGGGAG GAGGTGTCTC AGCTCCGGCA TGAGCTGTCC ATGAAGGATG
501 AGCTGCTTCA GTTCTACACC AGCGCAGCGG AGGAGAGTGA GCCCGAGTCC
551 GTTTGGCTCAA CCCCGTTGAA GAGGAATGAG TCGTCTCCTC CAGTCCAGAA
601 TTACTTTTCAT TTGGATTCTC TTCAAAGAA GCTGAAAGAC CTTGAAGAGG
651 AGAATGTTGT ACTTCGATCC GAGGCCAGCC AGCTGAAGAC AGAGACCATC
701 ACCTATGAGG AGAAGGAGCA GCAGCTGGTC AATGACTGCG TGAAGGAGCT
751 GAGGGATGCC AATGTCCAGA TTGCTAGTAT CTCAGAGGAA CTGGCCAAGA
801 AGACGGAAGA TGCTGCCCGC CAGCAAGAGG AGATCACACA CCTGCTATCG
851 CAAATAGTTG ATTTGCAGAA AAAGGCAAAA GCTTCCGCG TGGAAAATGA
901 AGAAGTTGTC CAGCATCTGG GGGCTGCTAA GGATGCCAG CGGCAGCTCA
951 CAGCCGAGCT CCGTGAGCTG GAGGACAAGT ACGCAGAGTG CATGGAGATG
1001 CTGCATGAGG CGCAGGAGGA GCTGAAGAAC CTCCGGAACA AAACCATGCC
1051 CAATACCACG TCTCGGCGCT ACCACTCACT GGGCCTGTTT CCCATGGATT
1101 CCTTGGCAGC AGAGATTGAG GGAACGATGC GCAAGGAGCT GCAGTTGGAA
1151 GAGGCCGAGT CTCCAGACAT CACTCACCAG AAGCGTGTCT TTGAGACAGT
1201 AAGAAACATC AACCAGGTTG TCAAGCAGAG ATCTCTGACC CCTTCTCCCA
1251 TGAACATCCC CGGCTCCAAC CAGTCTCCGG CCATGAACTC CCTCCTGTCC
1301 AGCTGCGTCA GCACCCCCCG GTCCAGCTTC TACGGCAGCG ACATAGGCAA
1351 CGTCGTCTCT GACAACAAGA CCAACAGCAT CATTCTGAA ACAGAGGCAG
1401 CCGACCTGGG AAACGATGAG CGGAGTAAGA AGCCGGGGAC GCCGGGCACC
1451 CCCAGGCTCC CACGACCTGG AGACGGCGCT GAGGCGGCTG TCCCTGCGCC
1501 GGGAGAACTA CCTCTCGGAG AGGAGGTTCT TTGAGGAGGA GCAAGAGAGG
1551 AAGCTCCAGG AGCTGGCGGA GAAGGGCGAG CTGCGCAGCG GCTCCCTCAC
1601 ACCCACTGAG AGCATCATGT CCCTGGGCAC GCCTCCCGC TTCTCCGAGT
1651 TCACCGGCTT CTCTGGCATG TCCTTCAGCA GCCGCTCCTA CCTGCCTGAG
1701 AAGCTCCAGA TCGTGAAGCC GCTGGAAGGT GATCACGCGG GGCCTCGGCC
1751 CCTCTCTGTC CTCTGGGGG ACTCCCTTTG GTCCCTGATC CACCTGCGGA
1801 AGGCGGGGCA CCTCTGTAC GCCTACTCCT TTTTCTTCCG CGACAGCCAC
1851 CCGCGCTGCT GGTTTGAGTT CCTCTGAGGG TGGTGCTCAG CCTAGGCCTC
1901 CGTCCCTCCC CTCTGGCTGG CAGGTGTGAC AATGCACACA TAGGCCATGA
1951 AACTCGCCGA GGAAGACAA GCATGTGCAC TGTGGTCTTC TAGTCTTTC
2001 CTTTGCCTTT AGAACCTTAG AAATAAAAC TTTTGTGGCG GTAGAGGCAC
2051 TGCTAACTGA TTCAAAAATT AATTAGGTTT TGCCGTGGG TGTGAGGAAT
2101 GCAGAAAATT AATGCTTTAG CTTTCTGCA GTTTTGGTGT CGGGGAGAGG
2151 TTCCAAGCAA ACTCTATTAA ATGGGGATT TTTTTCCTCC ATAACCACCT
2201 GAATGTGATT TGTGGGCTTA TGTGTTCTGA TTTGAACTTC ATATAGCAAG
2251 GTTGTGGCTT TTGGCAGATG CAGTATGTTG TGAGCGCGGC TCCTAGAGTC
2301 TACAATTTGG AGTCCAGGAA GGGGTGGCTG TGGAGACAAG TGAGTTTGTG
2351 ACCTCCGTAA GCCACCTTT TTCAGGGTCA GTTCATGTGT TAGTATCAGG
2401 GGCATCTCAG ATGATTAAAC TCATGGGAAA AACTTCCTCC TTCCCTCTCT
2451 CCCTCTTGCC CTCTGCTCTC TTTTTCCTTC TTTTTCCTTC AATTGGGGCA
2501 CTTATAAAAT GTTTTCCTTC TACCTGCTGC TACTCTGCCA AGAGCCACCA
2551 AGTGCTTATA TTTTTCATTT TTTACTCTTT TAGTTTGAA AGCCATATAC
2601 GTTTGAGAGG GTGTTTAAA ACTCTGTGTT ACACCTACGA TGCAAAGCCA
2651 AATCAGAACT TCTGTAAGGC AGAACTTTCC CAACTTTAAA AAAATTATTG
```

```

2701 TCCCCTCTAG GAGCCTTCTT AGACGTTTTT TCCTAATCAC CCCCCAAAGA
2751 CATTTTAATA CCACATATAT ATTGTTTATG TACTATATGT ATATACATAA
2801 ACAATACATA AGCAATACAT CTGTGGTATT AAAATTAAAA AGAATCCAAT
2851 TATGTTTACC TCAAAAGAAC CTGTTTTTGC TTCTTGGGAG CAATATTGCC
2901 CCTGTGAGAG TGCATGCTAT AAGGTAAGGT TGTGCTTGT AAAGACCCAA
2951 GACATGACTG GGTTCACAG TCTCCAAAGG AAGAGGGTGG GCTAGTTTGT
3001 TTTTATTATT ATTTTAAAT TGTATATTG GGGTCTTCT TAGAGTTCAG
3051 AAAAGGTATA GCTTACTCTT TTTTAATTGT TTATTTAGTT GTAAGCTTAG
3101 TGATTGTTT CTGATCCACA TTGTGTGTGT TCTTCAATAA AATCTTTCAT
3151 TTCTGCAATT TTAACAAAAA AAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 206 bp to 1531 bp; peptide length: 442

Category: similarity to known protein

Classification: unset

Prosites motifs: LEUCINE_ZIPPER (139-161)

```

1 MQKFIEADYY ELDWYEECS DVLCAERVGO MTKTYNDIDA VTRLLEEKER
51 DLELAARIGQ SLLKKNKTLT ERNELLEEQV EHIREEVSQ RHELSMKDEL
101 LQFYTSAAEE SEPESVCSTP LKRNESSSV QNYFHLDSLQ KKLKDLLEEN
151 VVLRSEASQL KTETITYEEK EQQLVNDVCK ELRDANVQIA SISEELAKKT
201 EDARQQUEI THLSQIVDL QKKAKACAVE NEELVQHLGA AKDAQRLTA
251 ELRELEDKYA ECEMELHEAQ EELKNLRNKT MPNTTSRRYH SLGLFPMDSL
301 AAEIEGTMRK ELQLEEAESP DITHQKRVFE TVRNINQVVK QRSLTSPMPN
351 IPGSNQSSAM NSLLSSCVST PRSSFYGS DI GNVVLDNKTN SIILETEAAD
401 LGNDERSKKP GTPGTPRLPR PGDGAEEAVP APGELPLGEE VL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfkd2_4c8, frame 2

PIR:S72555 huntingtin-associated protein HAP1 - human (fragment), N = 1, Score = 234, P = 8.6e-19

TREMBL:CEUT27A3 7 gene: "T27A3.1"; Caenorhabditis elegans cosmid T27A3., N = 1, Score = 226, P = 9.9e-16

PIR:S67495 huntingtin-associated protein HAP1-A - rat, N = 1, Score = 215, P = 1.6e-14

>PIR:S72555 huntingtin-associated protein HAP1 - human (fragment)
Length = 320

HSPs:

Score = 234 (35.1 bits), Expect = 8.6e-19, P = 8.6e-19
Identities = 66/189 (34%), Positives = 110/189 (58%)

```

Query: 109 EESEPEVCSTPLKRNE--SSSSVQNYFH---LDSLQKKLKDLEENVVLRSEASQLKTE 163
      EE+E + C+ P + S ++ + H L++LQ+KL+ LEEEN LR EASQL T
Sbjct: 28 EEAEEDLQCAHPCDAPKLISQEALLHQHCPQLEALQEKLRLLEENHQLREEASQLDT- 86

Query: 164 TITYEEKEQQLVNDVCKELRDANVQIASISEELAKKTEDAARQQEEITHLSQIVDLQKK 223
      E++EQ L+ +CV++ +A+ Q+A +SE L + E+ RQ+Q+E+ L +Q++ LQ++
Sbjct: 87 ---LEDEEQMLILECQVEQFSEASQMAELSEVLVLRLENYERQQQEVARLQAQVLKQQR 143

Query: 224 AKACAVENEELVQHLGAQKDAQRQLTAE--LRELEDKYAECME--MLHEAQEELKNL-RN 278
      + E E+L + L + K+ Q QL E L ++ AE + + + + + RN

```

Sbjct: 144 CRMYGAETEKLQKQLASEKEIQMQLQEEETLPGFQETLAEELRTSLRRMISDPVYFMERN 203
 Query: 279 KTMP--NTTSRRY 289
 MP +T+S RY
 Sbjct: 204 YEMPRGDTSSSLRY 216

Peptide information for frame 3

ORF from 1416 bp to 1874 bp; peptide length: 153
 Category: similarity to known protein
 Classification: unset

1 MSGVRSRGR APPGSHOLET ALRRLSLRRE NYLSERRFFE EEQERKLQEL
 51 AEKGLRSGS LPTESIMSL GTHSRFSEFT GFGMSFSSR SYLPEKLQIV
 101 KPLEGDHAGP RPLSVLLGDS LWSLIHLRKA GHLCHAYSFF FRDSHPRCWF
 151 EFL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_4c8, frame 3

TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds., N = 1, Score = 252, P = 5.5e-21

>TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds.
 Length = 469

HSPs:

Score = 252 (37.8 bits), Expect = 5.5e-21, P = 5.5e-21
 Identities = 57/98 (58%), Positives = 69/98 (70%)

Query: 8 GRRAPPGSHOLETALRRLSLRRNYLSERRFFEEQERKLQELAEKGLRSGSLTPTESI 67
 G+ P G DL TAL RLSLRR+NYLSE++FF EE +RK+Q LA++ E SG +TPTES+
 Sbjct: 27 GQGPSPGSDSLATALHRLSLRRQNYLSEKQFFAEWQRKIQLADQKEGVSGCVTPTESL 86
 Query: 68 MSLGTHSRFSEFTGFGMSFSSRSYLPEKLQIVKPLEG 105
 SL T SE T S S R ++PEKLQIVKPLEG
 Sbjct: 87 ASLCTTQ--SEITDLSSAS-CLRGFMPEKLQIVKPLEG 121

Pedant information for DKFZphfd2_4c8, frame 2

Report for DKFZphfd2_4c8.2

[LENGTH] 442
 [MW] 50020.14
 [pI] 4.77
 [HOMOL] TREMBL:AF040723_1 product: "neuroanl"; Homo sapiens neuroanl mRNA, complete cds. 5e-29
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-08
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YIL149c] 5e-08
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-08
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YIL138c] 6e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGR130c] 2e-07
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-06
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-06
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1643] 1e-06
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-06
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-06
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 4e-06
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 4e-06
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YNL250w] 2e-05
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 2e-05

```

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c]
5e-05
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]
1e-04
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 1e-04
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YNL272c] 3e-04
[FUNCAT] 08.16 extracellular transport [S. cerevisiae, YNL272c] 3e-04
[BLOCKS] BL01289B
[BLOCKS] BL00415M Synapsins proteins
[EC] 3.6.1.32 Myosin ATPase 2e-07
[PIRKW] tandem repeat 2e-07
[PIRKW] heterodimer 1e-06
[PIRKW] endocytosis 9e-07
[PIRKW] heart 1e-06
[PIRKW] transmembrane protein 4e-07
[PIRKW] zinc finger 9e-07
[PIRKW] metal binding 9e-07
[PIRKW] DNA binding 3e-06
[PIRKW] muscle contraction 2e-07
[PIRKW] acetylated amino end 3e-06
[PIRKW] actin binding 2e-07
[PIRKW] mitosis 1e-06
[PIRKW] microtubule binding 1e-06
[PIRKW] ATP 2e-07
[PIRKW] chromosomal protein 1e-06
[PIRKW] receptor 3e-08
[PIRKW] thick filament 2e-07
[PIRKW] phosphoprotein 8e-06
[PIRKW] glycoprotein 3e-08
[PIRKW] skeletal muscle 3e-06
[PIRKW] DNA condensation 1e-06
[PIRKW] alternative splicing 2e-06
[PIRKW] coiled coil 2e-07
[PIRKW] P-loop 2e-07
[PIRKW] heptad repeat 4e-07
[PIRKW] methylated amino acid 2e-07
[PIRKW] peripheral membrane protein 9e-07
[PIRKW] cardiac muscle 6e-06
[PIRKW] hydrolase 2e-07
[PIRKW] muscle 2e-06
[PIRKW] cytoskeleton 2e-06
[PIRKW] Golgi apparatus 4e-07
[PIRKW] calmodulin binding 9e-07
[SUPFAM] myosin motor domain homology 2e-07
[SUPFAM] tropomyosin TPM1 2e-06
[SUPFAM] giantin 4e-07
[SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
[SUPFAM] human early endosome antigen 1 9e-07
[SUPFAM] unassigned kinesin-related proteins 4e-07
[SUPFAM] M5 protein 8e-08
[SUPFAM] cytoskeletal keratin 3e-06
[SUPFAM] myosin heavy chain 2e-07
[SUPFAM] conserved hypothetical P115 protein 1e-06
[SUPFAM] centromere protein E 1e-06
[SUPFAM] pleckstrin repeat homology 2e-06
[SUPFAM] kinesin motor domain homology 4e-07
[PROSITE] LEUCINE_2IPPER 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 6.79 %
[KW] COILED_COIL 27.15 %

```

[illegible]

SEQ SLLKKNKTILTERNELLEEQVEHIREEVSQLRHELSMKDELLOFYTSAEESEPEPVCSTP
SEG
PRD hh
COILS cccccccccccccccccccccccccccccccccccccc.....

```

SEQ      LKRNESSSSVQNYFHLDSLQKKLKDLEENVVLRSEASQLKTETITYEEKEQQLVNDCKV
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```

[illegible]

Prosites for DKFZphfkd2 4c8.2

PS00029 139->161 LEUCINE ZIPPER PDOC00029

(No Pfam data available for DKFZphfkd2 4c8.2)

Pedant information for DKFZphfkd2.4c8, frame 3

Report for DKFZphfkd2 4c8.3

```
[LENGTH]      153
[MW]           17642.03
[pI]           9.38
[HOMOL]        TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens
mRNA for KIAA0549 protein, partial cds. 2e-12
[KW]           Alpha_Beta
[KW]           LOW COMPLEXITY      12.42 %
```

```
SEQ      MSGVRSRGRKAPPGSHDLETALRLSLRRNYLSERRFFEEEQERKLQLAEKGELRSGS
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccc
```



```
SEQ      LTPTESIMSLGTHSRFSEFTGFSGMSFSRRSYLPEKLQIVKPLEGDHAGRPRLSVLLGDS
SEG      .....
PRD      cccccceeeccccceeeccccccccccccccccchhhhhhccccccccccccceeeeccc
```



```
SEQ      LWSLIHLRKAGHLCHAYSFFFRDSPRCWFEFL
SEG      .....
PRD      chhhhhhhhhccccccccceeecccccccccccc
```

(No Prosite data available for DKFZphfkd2 4c8.3)

(No Pfam data available for DKFZphfk2 4c8.3)

DKFZphfkd2_4k14

group: intracellular transport and trafficking

DKFZphfkd2_4k14.3 encodes a novel 254 amino acid putative GTP-binding protein nearly identical to Rab6.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory (biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes.

rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport.

The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

strong similarity to Rab6

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3084 bp

Poly A stretch at pos. 3061, polyadenylation signal at pos. 3043

```
1 GGGGCACTCA GCAGGTTGGG CTGCGGCGGC GCGGCTGGG GAAGCCGAAG
51 CGCCGCGCGT GAGAGATCCC GGATACATCT GCGGTTTGGG CTCCGCCACC
101 CTCCGTCTCT CTCCCGCAGG TCTCTGAGCC GGTGCGGAA GAGGGGAACG
151 GCCCTAGCCT TGGGAAGCCA AAGCACACCC CTGGCTCCCG CCGACACCGC
201 CCTCCTTCCC TTCCAGCCG CGGGCCTCGC TCCGTGCTCG GCTACTCTGC
251 CGGAGGCGCG CGGCGGCTGC CAGTCTGTGG CGAGCCCTGC TGCCCTCCAG
301 CCGGCTTCT CCAGCCGGGC TCCTCCACCG GCCCTTGCAG GGGCACAGAG
351 AGCTCGGCGC CCGCCCTTCC GCTCGCCTTT TTCGTCAGCC GGCTGGAGGA
401 GCATCGGTCC GGGAGGTCTC TGGGCTGAGG CGGCGACAGC TCCTCTAGTT
451 CCACCATGTC CGCGGCGGGA GACTTCGGGA ATCCGCTGAG GAAATCAAG
501 CTGGTGTTCC TGGGGGAGCA AAGCGTTGCA AAGACATCTT TGATCACCAG
551 ATTCAGGTAT GACAGTTTTC ACAACACCTA TCAGGCAATA ATTGGCATTG
601 ACTTTTATC AAAAATATG TACTTGGAGG ATGGAACAAT CGGGCTTCGG
651 CTGTGGGATA CGGCGGGTCA GGAACGTCTC CGTAGCCTCA TTCCAGGTA
701 CATCCGTGAT TCTGCTGCAG CTGTAGTAGT TTACGATATC ACAAATGTTA
751 ACTCATTTCA GCAAACTACA AAGTGGATTG ATGATGTCAG AACAGAAAGA
801 GGAAGTGATG TTATCATCAC GCTAGTAGGA AATAGAACAG ATCTTGCTGA
851 CAAGAGGCAA GTGTCAGTTG AGGAGGGAGA GAGGAAAGCC AAAGGGCTGA
901 ATGTTACGTT TATTGAAACT AGGGCAAAAA CTGGATACAA TGTAAAGCAG
951 CTCTTTCGAC GTGTAGCAGC AGCTTTGCCG GGAATGGAAG GCACACAGGA
1001 CGGAGACAGA GAAGACATGA GTGACATAAA ACTGGAAGAG CCTCAGGAGC
1051 AAACAGTCAG CGAAGGGGGT TGTTCCTGCT ACTCTCCCAT GTCATCTTCA
1101 ACCCTCTCTC AGAAGCCCCC TTACTCTTTC ATTGACTGCA GTGTGAATAT
1151 TGGCTTGAAC CTTTCCCTT CATTAAATAC GTTTGCAAT TCATCATTCG
1201 TGCCTGTCTC GTGGAGGTGA TCTATTAGCT TCACAAGCAC AAAAAAGTC
1251 AGCGTCTTCA TTATTATAT TTACAAAAA GCCAAATTAT TTCAGCATAT
1301 TCCGCTGATA ACTTTAAAAA TTAGATACAT TTCTTAACA TTTTCTCTT
1351 TTTTAATGTT ATGATAATGT ACTTCAAAAT GATGGAAATC TCAACAGTAT
1401 GAGTATGGCT TGGTTAACGA GCAGTATGTT CACAGCCTGC TTTATCTCTC
1451 CTTGCTCTTC TCACCTCTCC CTTACCCCGT TCCCTATTTC CGTGTCTTCA
1501 CCTAGCCTCC CCCCCTTCC TCAAAACAAA CAAGAGATGG CAAAGCAGCA
1551 GTCCGACCAA GCCCCTGGA ATTATCCTTT AATTTTACAG ATACCACTTG
1601 CTGTAGGCTG TGGACCAAGA TGTCCAGAAT TATTCTTGAG CACTGATGTA
1651 AATTACTTAG ATCTTCTTTG AGGTGAGAA TCAAGCATCA CGGTAGGCAG
1701 TGCTTGAATG AGAAAAGCCT CCTGGTGCAAT CTTCAAAATG AGTCCTAAAG
1751 AACATACTGA GTACTTATAA GTAGCAGAAC ATAAAAATGTA TTTCTGACTA
1801 ACACAAATGG TCCTTTCACA TGTGCTTTAT TAGACTCTGG GAGAGAAAAAG
1851 TAACCAAGTG CTTCAAGACA GGTTTTATAG ATTTACTTCT TCATGGTAAG
1901 ATAATGAAGT TCTAATGAAC TATTCTCTCC AAGGTTTTAA AATTGTCAAG
1951 AGTTATCTCG TTTGTTTAAA AAGTAAGAAA CCTCTGTAAG CAATAGATT
2001 TGCTTGGGTT TTCTTCTTAA AAAAAATAAT ACTATGCAGG CAAGACACCA
2051 TAAAAGTTTA ATTCCTTACA GAAGAACCAG TGGAAGAATT TAAATTTGGC
2101 ACTACGATCA AAACACTGTA ATTAGCAGAA ATAACGATAT CTAAGGCTTA
2151 CCAGCAAAAG AACCTCAGC AGAATAGCAA AAACCTTGCT CAGGACATT
2201 GAGGTCAAAAT TGAAGACGGA AGACGGAAC CGGAACCGT TTTCTGTAA
2251 GCCCTAGAG GCAGATCAGG TAAGCATACA TAGTAGAGGG AAAGGAGAGA
2301 ATGGAAATAA AACTGAATAT TATGCAGATT TATGCCTTAT TTTTAGCAT
2351 TTTTAAAGGT TGGGTCTTTC AGGCTGGTTT TGGTTGTAT TAGATCTGTA
2401 TAGTTTAGTG ATTTAGTTT ATATTAAAGC TACGATTAAT ATTTTCTTT
2451 TGGGATATT TCTTGTCTT TTTTAAAAA CAACTTTCCA TTTTAGATG
```

```

2501 TTTCGTTGAA TCTATTTAGA GCTTCACCAT GGCAATATGT ATTTCCCTTA
2551 AAACACTGCA AACAAATATA CTAGGAGTGT GCCCTTTTAA TCTTTACTAG
2601 TTATTGTGAG ACTGCTGTGT AAGCTAATAA ACACATTGTG AAAAACATTG
2651 TTTGCAGGAA GAAAACTTCG AGTTACAGGT CAGGAAAAGC CTGCTGAATT
2701 TATGTTGTAA ACGTTACTTA ACACAGTATA AAGATGAAAA GACAACAAAA
2751 GTATCTTCAT ACTTCCTCAT CCCCTCATTT CAACAAAACC TTAAACTGGG
2801 AGAACCTTAG TCCCCTCTCT TTCCTCTTCC TCCTCCACTT CCCACTTATT
2851 GCCACTTTGT AATATTCAGA GAGCACTTGG ATTATGGATC TGAATAGAGA
2901 AATGCTTACA GATAATCATT AGCCACATA CCAGTAACTT ATACTTAAAG
2951 ATGGGATGGA GTTATAAGT GCTTTTATAA TCCAATATAA TTGCTAAAGG
3001 CAAGGGTTGA CTCCTTGTGT TATTTTGACA TGGCATGTCC TGAATAAAT
3051 ATTGGTTCAC TATGAAAAA AAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

98382468:
Rab proteins.

97203146:
GTP-bound forms of rab6 induce the redistribution of Golgi
proteins into the endoplasmic reticulum.

Peptide information for frame 3

ORF from 456 bp to 1217 bp; peptide length: 254
Category: strong similarity to known protein
Classification: unset
Prosites motifs: BACTERIAL_OPSIN_RET (45-57)

```

1 MSAGGDFGNP LRKFKLVFLG EQSVAKTSLI TRFRYDSFDN TYQAIIGIDF
51 LSKTMYLEDG TIGLRLWDTA GQERLRLSLIP RYIRDSAAAV VVYDITNVNS
101 FQQTWKIDD VRTERGSDVI ITLVGNRTDL ADKRQVSVEE GERKAKGLNV
151 TFIETRAKTG YNVKQLFRRV AAALPGMEST QDGSREDMSD IKLEKPQEQT
201 VSEGGCSCSY PMSSSTLPQK PPYSFIDCSV NIGLNLFPST ITFCNSSLPL
251 VSWR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_4k14, frame 3

PIR:G34323 GTP-binding protein Rab6 - human, N = 1, Score = 944, P = 6.5e-95

TREMBL:CET25G12.2 gene: "T25G12.4"; Caenorhabditis elegans cosmid T25G12., N = 1, Score = 756, P = 5.4e-75

TREMBL:NTNTRAF_1 gene: "Nt-rab6"; Nicotiana tabacum SR1 Nt-rab6 mRNA, complete cds., N = 1, Score = 698, P = 7.6e-69

TREMBL:D84314_1 product: "rab6"; Drosophila melanogaster mRNA for rab6, complete cds., N = 1, Score = 836, P = 1.9e-83

PIR:T01588 small GTP-binding protein F16B22.10 - Arabidopsis thaliana, N = 1, Score = 704, P = 1.8e-69

>PIR:G34323 GTP-binding protein Rab6 - human
Length = 208

HSPs:

Score = 944 (141.6 bits), Expect = 6.5e-95, P = 6.5e-95
Identities = 186/208 (89%), Positives = 190/208 (91%)

Query: 1 MSAGGDFGNPLRKFKLVFLGEQSVAKTSLITRFYDSFDNTYQAIIGIDFLSKTMYLEDG 60
MS GGDGFGNPLRKFKLVFLGEQSV KTSLITRF YDSFDNTYQA IGIDFLSKTMYLED
Sbjct: 1 MSTGGDFGNPLRKFKLVFLGEQSVGKTSLITRFMYDSFDNTYQATIGIDFLSKTMYLED 60

Query: 61 TIGLRLWDTAGQERLRLSLIPYIRDSAAAVVYDITNVNSFQQTTKWIDDVTERGSDVI 120
T+ L+LWDTAGQER RSLIP YIRDS AVVVYDITNVNSFQQTTKWIDDVTERGSDVI
Sbjct: 61 TVRLQLWDTAGQERFRLSLIPYIRDSVAVVVYDITNVNSFQQTTKWIDDVTERGSDVI 120

Query: 121 ITLVGNRTDLADKRQVSVEEGERKAKGLNVFTIETRAKTGYNVKQLFRRVAAALPGMEST 180
I LVGN+TDLADKRQVS+EEGERKAK LNV FIET AK GYNVQLFRRVAAALPGMEST
Sbjct: 121 IMLVGNKTDLADKRQVSIEEGERKAKELNVFIETSAKAGYNVQLFRRVAAALPGMEST 180

Query: 181 QDGSREDMSDIKLEKQPQEQTVSEGGCSC 208
QD SREDM DIKLEKQPQEQ VSEGGCSC
Sbjct: 181 QDRSREDMIDIKLEKQPQEQPVSEGGCSC 208

Pedant information for DKFZphkd2_4k14, frame 3

Report for DKFZphkd2_4k14.3

[LENGTH] 254
[MW] 28385.29
[pI] 7.58
[HOMOL] PIR:G34323 GTP-binding protein Rab6 - human 1e-102
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YLR262c]
7e-60
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YLR262c] 7e-60
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YOR089c] 2e-33
[FUNCAT] 08.19 cellular import [S. cerevisiae, YOR089c] 2e-33
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YOR089c] 2e-33
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YOR089c]
2e-33
[FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
YGL210w] 3e-28
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 8e-27
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w]
8e-27
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YOR101w]
2e-21
[FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 2e-21
[FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YOR101w]
2e-21
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 2e-21
[FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S.
cerevisiae, YOR101w] 2e-21
[FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YOR101w] 2e-21
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 6e-19
[FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 6e-19
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 6e-19
[FUNCAT] 04.07 rna transport [S. cerevisiae, YOR185c] 6e-16
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR185c] 6e-16
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YOR185c] 6e-16
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 4e-13
[FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 4e-13
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 2e-09
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-08
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YLR229c] 8e-08
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL180c] 1e-05
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR094w] 5e-05
[BLOCKS] BL01115A GTP-binding nuclear protein ran proteins
[SCOP] dlas3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domai 1e-32
[SCOP] dlmh1_ 3.29.1.4.2 Rac1 [Human (Homo sapiens)] 2e-51
[SCOP] d5p21_ 3.29.1.4.1 cH-p21 Ras protein [human (Homo sapiens)] 7e-53
[SCOP] dlhura_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1) [human (Homo sapiens)] 1e-46
[SCOP] dla2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2) [Do 6e-60
nucleus 2e-14
[PIRKW] cell cycle control 5e-15
[PIRKW] membrane trafficking 3e-71
[PIRKW] endoplasmic reticulum 1e-29
[PIRKW] phosphoprotein 1e-29
[PIRKW] prenylated cysteine 2e-36
[PIRKW] signal transduction 5e-15
[PIRKW] transforming protein 5e-30
[PIRKW] purine nucleotide binding 1e-28
[PIRKW] alternative splicing 1e-18
[PIRKW] P-loop 3e-71

[PIRKW] lipoprotein 2e-36
 [PIRKW] proto-oncogene 1e-20
 [PIRKW] methylated carboxyl end 1e-20
 [PIRKW] membrane protein 1e-29
 [PIRKW] GTP binding 3e-71
 [PIRKW] thiolester bond 1e-29
 [PIRKW] Golgi apparatus 1e-29
 [SUPFAM] ras transforming protein 1e-76
 [PROSITE] BACTERIAL_OPSIN_RET 1
 [PFAM] Ras family (contains ATP/GTP binding P-loop)
 [KW] Alpha_Beta
 [KW] 3D

SEQ MSAGGDFGNPLRKFKLVFLGGEQSVAKTSLITRFYDSFDNTYQAIIGIDFLSKTMYLEDG
 lkao-CCEEEEEEECTTTTCHHHHHHHHHHCCCCCTTTC-EEEEEEEETTE

SEQ TIGLRLWDTAGQERLRSIPRYIRDSAAAVVYDITNVNSFQQTTKWIDDVTERGSDVI
 lkao- EEEEEEEECTTTTCHHHHHHHHHHCCCCCEEEEEETTTTHHHHHHHHHHHHHHHHTTCC

SEQ ITLVGNRTDLADKRQVSVEEGERKAKGLNVTFIETRAKTGYNVKQLFRRVAAALPGMEST
 lkao- EEEEEETTTGGGCCCHHHHHHHHHHCCCCCEECTTTTHHHHHHHHHHH.....

SEQ QDGSREDMSDIKLEKPQEQTVSEGGCSCYSPMSSSTLPQKPPYSFIDCSVNIGLNLFPSSL
 lkao-

SEQ ITFCNSSLLPVSWR
 lkao-

Prosites for DKF2phfkd2_4k14.3

PS00327 45->57 BACTERIAL_OPSIN_RET PDOC00291

Pfam for DKF2phfkd2_4k14.3

HMM_NAME Ras family (contains ATP/GTP binding P-loop)

HMM *KLVLI GDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK
 KLV++G+ +V K++L RF +++F++ Y + IG+DF++KT++++ TI
 Query 15 KLVFLGEQSVAKTSLITRFYDSFDNTYQAIIGIDFLSKTMYLEDGTIG 63
 HMM LQIWDTAGQERYRSMRPYYRGAMGFMVYDITNRqSFENIrNWweEIrR
 L +WDTAGQER RS+ P Y+R++ ++++VYDITN SF+ ++W++++R+
 Query 64 LRLWDTAGQERLRSIPRYIRDSAAAVVYDITNVNSFQQTTKWIDDVRT 113
 HMM HCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKTN
 + ++V+I LVGN +DL+D+RQVS EEG+ A+ ++ + F+ET AKT+
 Query 114 ERG--SDVIITLVGNRTDLADKRQVSVEEGERKAKGLN-VTFIETRAKTG 160
 HMM iNVEEAFMEIvReIlqrMqe.q.NgteNinidQpsrnrk...rCCCIM*
 +NV++ F +++ +++ +++ + ++++++I+ +++ + +C+ +
 Query 161 YNVKQLFRRVAAALPGMESTQDGSREDMSDIKLEKPQEQTVSEGGCS-C 208

DKFZphfkd2_4m11

group: transmembrane protein

DKFZphfbr2-4m11 encodes a novel 159 amino acid protein with weak similarity to the putative membrane protein YMR034c of *S. cerevisiae*.

The novel protein contains 4 transmembrane regions.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker of neuronal cells.

weak similarity to YMR034c

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1749 bp

Poly A stretch at pos. 1727, polyadenylation signal at pos. 1713

```
1 GGGGTCCTCA AAGCCGCCGG AGCAACCCCC AGGTCTTTAC TTTACAATCG
51 GCAATTTGAC TTGCTCTGCT GCATGTCTGG AGGGACCAAG GAAAGTGTGG
101 AGACGCTCCA AGGATTAGGT GATCGGAGCT TGAAGAAGAA AAAAGCCAAA
151 CAAATAAACA AAACCCACCC ACCCTAACGA ATATGAGGCT GCTGGAGAGA
201 ATGAGGAAAG ACTGGTTCAT GGTCCGAATA GTGCTGGCGA TCGCTGGAGC
251 TAACTGGAG CCGTCCATAG GGTGAATGG GGGACCACTG AAGCCAGAAA
301 TAACGTATC CTACATTGCT GTTGCAACAA TATTCTTTAA CAGTGGACTA
351 TCATTGAAAA CAGAGGAGCT GACCACTGCT TTGGTGCATC TAAAACTGCA
401 TCTTTTATT CAGATCTTTA CTCTTGCAAT CTCCCAAGCA ACAATATGGC
451 TTTTCTTCA GCTTTTATCA ATCACACCCA TCAACGAATG GCTTTTAAAA
501 GGTTTGCAAG CAGTAGGTTG CATGCCTCCG CCTGTGTCTT CTGCAGTGAT
551 TTTAACCAAG GCAGTTGGTG GAAATGAGGC AGCTGCAATA TTTAATTCAG
601 CCTTTGGAAG TTTTGTGTA AGTAACATA GTTAACTTG TCTATTACAA
651 CTTTGTGCTG GATATTGTGT ATATGAAAGA TTTAGTGAAA GCTGGATTTG
701 TTTTACTCTT TGGTTAAGTA TAAAAATTGT TGAATCTTTT CATGTGCCAG
751 TATCCATACC CTGAAGAAAA GTAGTTAATG AATAAGCAA ATGTTCTCTT
801 ACAATATATT TTGGAGGTTT GGATTTTAAA ATTCCATTTA ATGAATTCAA
851 GGAATCAATT AAAACACTAT GTGTCTCCTT ATAGAGGTTA TGTCAATATA
901 TTGATCAATT AATGAGGTCT TTTAGATTAT TATTATTTTG TATCATGGGA
951 CTGAGGATTT TGAAGAGGAA ACATGACCCA GCTGGTCAGA AAGGGAATGC
1001 TAATTACTTT GTTGACATGC CATTATTTT GTACATTTCA CTGTCAAAGA
1051 AGCTACTGGC TTGGATGCTT CTGAGAAATC TATGTGAGAA AAAATTGAA
1101 AGGAAGATAT GACTAATGAG TAATTGCAA GTAAATGTTG TATCTATATA
1151 TATATATATA TAAAGATTCA AAGTAGTTC AGCTTTCATA AGTAGAACCA
1201 ATATAAGGAC GTTGTTTTAG CATTTTTAA TATTATTTT AAATAAATGA
1251 TGTAACAGAG GCTTGATTTG TGTATGAAA GATTGAGAAA CTAATTTTTC
1301 TGTTGATTTA ATTTTGTGT GCCTTAAAC TTTGTTAAAT TCCTGAAGTT
1351 AATTACTATA TTGTACTTTT TGGGGCATAA CTCATTAGCA GATATGTAGT
1401 GCAGTGATTT ACAAATAATT GAGAGTAAA TCAGTGATGT ATAAACTAGT
1451 TCATGAGTCT AGGTAAAATA TCAATTACCT CTGTTTAAAA TGCTCTGTTA
1501 ATTATTATG TATGTATTTA AATGTAGTTA AAGCTTTTAA ACATGTTGTT
1551 ACATAGTGTT AATTCTACAC AGTGCTACAC AGCTTTTAGT GTCACATAGC
1601 CTTACAGAGT TTATAATGAT GTAGCATCTG CAAAATATAT GCATAGCTTA
1651 TATCTATTTT TTATAGAGCC AGTAATGGTT TTTGTGATGC TGTATTACTT
1701 CTGGGTTTTA GACARTAAAG TCTGTTTAA CAAAAAATAA AAAAAAATAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

```

1  MRLLERMIRD  WFMVGIVLAI  AGAKLEPSIG  VNGGPLKPEI  TVSYIAVATI
51  FFNLSGLSKT  EELTSALVHL  KLHLFIQIFT  LAFFPATWL   FLQLLSITPI
101 NEWLLKGLQT  VGCMPPPVSS  AVILTKAVGG  NEAAAFNSA   FGSFLVSKHS
151 LGLLOLLL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2 4m11, frame 3

PIR:A65015 yfeH protein - Escherichia coli (strain K-12), N = 1, Score = 131, P = 4.2e-08

>PIR:S53951 probable membrane protein YMR034c - yeast (*Saccharomyces cerevisiae*)
Length = 434

HSPs:

Score = 171 (25.7 bits), Expect = 3.2e-12, P = 3.2e-12
Identities = 38/144 (26%), Positives = 72/144 (50%)

```

Query:      5  ERMRKDFWFMVGIVLAIAGAKLEPSIGVNGGGLPKPEITVSYIAVATIFFNSGLSLKTEELT 64
            E ++ WF ++ ++ I A+ P+ +GG +K + ++ Y VA IF SGL +K+ L
Sbjct:     18  EFLKSWQFFICLAILLVIARFAPNFARDGGGLIKGQYSIGYGCVAWIFLQSGLGKMSRSLM 77

Query:     65  SALVHLKLHLFIQIFTLAFPPATIWLF---LQLSITPINEWLLKGLQTVGCMPPPVSSA 121
            + +++ + H I + + + ++ F ++ + I +W+L GL P V+S
Sbjct:     78  ANMLNWRHAHATILVLSFLITSSIVYGCCAVKAANDPKIDDVVLIGLILTATCPTTVASN 137

Query:    122  VILT KAVGGNEAAAIFNSAFGSFL 145
            VI+T GGN + G+ L
Sbjct:    138  VIMTTNAGGNSLLCVCEVFIGNLL 161

```

Pedant information for DKFZphfkd2 4m11, frame 3

Report for DKFZphfkd2 4m11.3

```
[LENGTH]      159
[MW]           17282.92
[pI]           9.06
[HOMOL]        PIR:S53951 probable membrane protein YMR034c - yeast (Saccharomyces cerevisiae)
5e-12
[FUNCAT]       99 unclassified proteins           [S. cerevisiae, YMR034c] 2e-13
[PROSITE]      MYRISTYL 2
[PROSITE]      PKC_PHOSPHO_SITE 1
[KW]           TRANSMEMBRANE 4
```

```

SEQ      MRLLEMRKQDFMVGIVLAIAGAKLEPSIGVNGGPKLPEITVSYIAVATIFFNSGLSLKT
PRD      cccchhhhhhhhhhhhhhhhhhhhhcccccceccccccccccccccccchhhh
MEM      ....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      EELTSLVHLKHLFIQIFTLAFFPATIWLFLQLLSITFINEWLLKGLQTVCMPPPVSS
PRD      hhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhcccccchhhhhhhheeececccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      AVILTQAVGGNEAAAFNSAFGSLVSKHSLTCLQLLL
PRD      cecececcccchhhhhhhccccccececececececcc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

```

Prosite for DKFZphfk2 4m11.3

PS00005	57->60	PKC_PHOSPHO_SITE	PDOC00005
PS00008	15->21	MYRISTYL	PDOC00008
PS00008	129->135	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfd2 4m11.3)

PAGE INTENTIONALLY LEFT BLANK

DKFZphutel_17k7

group: uterus derived

DKFZphutel_17k7 encodes a novel 520 amino acid protein with weak similarity to *S. Cerevisiae* Fip1.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to *S.cerevisiae* Fip1

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1914 bp

Poly A stretch at pos. 1897, polyadenylation signal at pos. 1867

```
1 CGGACGCGTG GCGGACGCG TGGGGCCTTC CTGGGATTGG AGTCTCGAGC
51 TTTCTTCGTT CGTTCGCCGG CGGGTTCGCG CCCTTCTCGC GCCTCGGGGC
101 TGCGAGGCTG GGGAGGGGT TGGAGGGGCG TGTTGATCGC CGCGTTAAG
151 TTGCGCTCGG GCGGCGCATG TCGGCGGCG AGGTCGAGCG CCTAGTGTCTG
201 GAGCTGAGCG GCGGACCGG AGGGGATGAG GAGGAAGAGT GGCTCTATGG
251 CGATGAAAAT GAAAGTTGAAA GGCCAGAAGA AGAAAATGCC AGTGCTAATC
301 CTCCATCTGG AATTGAAAGT GAAACTGCTG AAAATGGTGT ACCAAAACCG
351 AAAGTGACTG AGACCGAAGA TGATAGTGAT AGTGACAGCG ATGATGATGA
401 AGATGATGTT CATGTCACCTA TAGGAGACAT TAAAACGGGA GCACCACAGT
451 ATGGGAGTTA TGGTACAGCA CCTGTAAATC TTAACATCAA GACAGGGGGA
501 AGAGTTTATG GAACTACAGG GACAAAAGTC AAAGGAGTAG ACCTTGATGC
551 ACCTGGAAGC ATTAATGGAG TTCCACTCTT AGAGGTAGAT TTGGATTCTT
601 TTGAAGATAA ACCATGGCGT AAACCTGGTG CTGATCTTTC TGATTATTTT
651 AATTATGGGT TTAATGAAGA TACCTGGAAA GCTTACTGTG AAAAAACAAA
701 GAGGATACGA ATGGGACTTG AAGTTATACC AGTAACCTCT ACTACAAATA
751 AAATTACGGT ACAGCAGGGA AGAAGTGGAA ACTCAGAGAA AGAAACTGCC
801 CTTCCATCTA CAAAAGCTGA GTTTACTTCT CCTCCTTCTT TGTTCAAGAC
851 TGGGCTTCCA CCGAGCAGGA GATTACCTGG GGCAATTGAT GTTATCGGTC
901 AGACTATAAC TATCAGCCGA GTAGAAGGCA GGCGACGGGC AAATGAGAAC
951 AGCAACATAC AGGTCCTTTC TGAAAGATCT GCTACTGAAG TAGACAACAA
1001 TTTTAGCAAA CCACCTCCGT TTTTCCCTCC AGGAGCTCCT CCCACTCACC
1051 TTCCACCTCC TCCATTTCTT CCACCTCCTC CGACTGTCAG CACTGCTCCA
1101 CCTCTGATTC CACCACCGGG TTTTCTCTCT CCACCAGGCG CTCCACCTCC
1151 ATCTCTTATA CCAACAATAG AAAGTGGACA TTCCTCTGGT TATGATAGTC
1201 GTTCTGCACG TGCATTTCCA TATGGCAATG TTGCCTTTCC CCATCTTCTT
1251 GGTTCGTGCTC CTTCTGGGCC TAGTCTTGTTG GACACCAGCA AGCAGTGGGA
1301 CTATTATGCC AGAAGAGAGA AAGACCGAGA TAGAGAGAGA GACAGAGACA
1351 GAGAGCGAGA CCGTGATCGG GACAGAGAAA GAGAACGCAC CAGAGAGAGA
1401 GAGAGGGAGC GTGATCACAG TCCTACACCA AGTGTTTTCA ACACGGATGA
1451 AGAACGATAC AGATACAGGG AATATGCAGA AAGAGGTTAT GAGCGTCACA
1501 GAGCAAGTCG AGAAAAAGAA GAACGACATA GAGAAAGACG ACACAGGGAG
1551 AAAGAGGAAA CCAGACATAA GTCTTCTCGA AGTAATAGTA GACGTCGCCA
1601 TGAAAGTGAA GAAGGAGATA GTCACAGGAG ACACAAACAC AAAAAATCTA
1651 AAAAGAAGCA AGAAGGAAAA GAAGCGGGCA GTGAGCCTGC CCCTGAACAG
1701 GAGAGCACCG AAGCTACACC TGCAGAATAG GCATGGTTTT GGCCTTTTGT
1751 GTATATTAGT ACCAGAAGTA GATACTATAA ATCTTGTTAT TTTTCTGGAT
1801 AATGTTTAAAG AAATTTACCT TAAATCTTGT TCTGTTTGTT AGTATGAAAA
1851 GTTAACTTTT TTTCCAAAAT AAAAGAGTGA ATTTTTCATG TTAAGTTAAA
1901 AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 168 bp to 1727 bp; peptide length: 520
Category: similarity to known protein

```

1 MSAGEVERLV SELSGGTGGD EEEEWLYGDE NEVERPEEEN ASANPPSGIE
51 DETAENGVPK PKVTETEDDS DSDSDDDEDD VHVITGDIKT GAPQYGSYGT
101 APVNLNIKTG GRVYGTGTGK VKGVLDLAPG SINGVPLLEV DLDSFEDKPW
151 RKPGADLSDY FNYGFNEDTW KAYCEKQKRI RMGLEVIPVT STTNKITVQQ
201 GRTGNSEKET ALPSTKAFT SPPSLFKTGL PPSRRLPGAI DVIQQTITIS
251 RVEGRRRANE NSNIQVLSEK SATEVDNNEF KPPFFPPGA PPTHLPPIPF
301 LPPPTVSTA PPLIPPPGFP PPPGAPPSL IPTIESGHSS GYDSRSARAF
351 PYGNVAFPHL PGSAPSWPSL VDTSKQWYY ARREKDRRE RDRDRERDRD
401 RDRERERTRE RERERDHSPT PSVFNSDEER YRYREYAERG YERHRASREK
451 EERHRERRHR EKEETRHKSS RSNSRRRHES EEGDSHRRHK HKKSKRSKEG
501 KEAGSEPAPE QESTEATPAE

```

BLASTP hits

Entry AF016427.4 from database TREMBL:
gene: "F32D1.9"; Caenorhabditis elegans cosmid F32D1.
Score = 392, P = 1.8e-36, identities = 156/519, positives = 212/519

Entry S62454 from database PIR:
hypothetical protein SPAC22G7.10 - fission yeast (Schizosaccharomyces pombe)
Score = 246, P = 2.0e-22, identities = 62/163, positives = 91/163

Entry A56545 from database PIR:
FIP1 protein - yeast (Saccharomyces cerevisiae)
Score = 186, P = 2.9e-16, identities = 56/206, positives = 92/206

Alert BLASTP hits for DKFZphutell_17k7, frame 3

TREMBLNEW:AF109907.1 product: "S164"; Homo sapiens S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds., N = 2, Score = 236, P = 1.5e-16

>TREMBLNEW:AF109907.1 product: "S164"; Homo sapiens S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds.

Length = 735

HSPs:

Score = 236 (35.4 bits), Expect = 1.5e-16, Sum P(2) = 1.5e-16
Identities = 51/120 (42%), Positives = 76/120 (63%)

```

Query: 383 REKDRDRERDRDRDRDRDRERERTREERERERDHSPTPSVFNSDEERYRYREYA---ER 439
      REK+++RER+R+R+RDRDR +ER+R R+RER+RD S + +++R R RE + ER
Sbjct: 227 REKEKERERERERDRDRDRDKERDRDRDRDRDRDRERSS-DRNKDRSRSREKSRDRER 285

```

```

Query: 440 GYERHRASREKEERHRER-RHREKEETRHKSSRSNSRRHSEEGDSHRRHKHKKSKRSK 498
      ER R + ER RER R RE+E R + + +R E +E D++ R K ++ R K
Sbjct: 286 ERERERERERERERERERERERERERERERERERERERERERERERERERERERERERER 345

```

Query: 499 E 499

E

Sbjct: 346 E 346

Score = 214 (32.1 bits), Expect = 4.4e-14, Sum P(2) = 4.4e-14
Identities = 50/133 (37%), Positives = 75/133 (56%)

```

Query: 383 REKDRDR-ERDRDRERDRDRDRERERTREERERERDHSPTPSVFNS-DEERYRYREYAERG 440
      RE++R+R ER+R+RER+R+R+++ER RERER+RD T D ER R R+ ER
Sbjct: 208 REREREREREREREREREREREKEKERERERERDRDRDRDKERDRDRDRDRDRDR-RERS 266

```

```

Query: 441 YERHRASREKEERHRERHRREKEETRHKSSRSNSRRHSEEGDSHRRHKHKKSKRSKEG 500
      +R++ E+ R+R RE+E R + R R R E + R + ++ K K
Sbjct: 267 SDRNKDRSRSREKSRDRE-RERERERERE-RERERERERERERERERERERERERERERERER 324

```

Query: 501 KEAGSEPAPEQESTE 515

+E E A E+ E

Sbjct: 325 REEDEEDAYERKKLE 339

Query: 443 RHRASREK-----EERHRRRRHR---EKEETHRKSSRSNSRRRRHES-EEGDSHRRH-KH 491
 + A +E+ ER + R + E+EE R + ++R E E+ D R K+
 Sbict: 345 KEAAQERLKNWEIERKRTREYEKEAEEREERRREMAKRLKEFLDYDDDRDDPKY 404

Query: 492 -----KKSRSKEGKEAGSEPAPEQESTE 515
+K R +E + E ++E E
Sbjct: 405 YRGSAQKRLRDREKEMEADERDRKREKEE 434

Score = 162 (24.3 bits), Expect = 2.4e-08, Sum P(2) = 2.4e-08
Identities = 45/141 (31%), Positives = 74/141 (52%)

Query: 372 DTSKQWDYYARREKDRDRDRDRERDRDRERERTREERERERDHSPTPSVFNSEERY 431
+ SK D + + E+++ ++ +E +++R RERER RERERER + ER
Sbjct: 172 EISKFRDTHKKLEEEKGKKEKERQEIEKER-RERERERERERERRERERER--ERERERE 228

Query: 432 RYREYAERGYERHRASREKEERHRER-RHREKEETHKSSRSNSRRRHSEEGDSHRRHK 490
+ +E ER ER R +ER R+R R R+++ R +SS N R E+ R +
Sbjct: 229 KEKE-RERERERDRDRRTKERDRDRDRERDRDRDRERSSDRNKDRSRSEKSRDRERER 287

Query: 491 HKKSRSKEGKEAGSEPAPEQE 512
++ +R +E +E E E+E
Sbjct: 288 ERERERERE-RERERERERE 308

Score = 137 (20.6 bits), Expect = 1.2e-05, Sum P(2) = 1.2e-05
Identities = 48/152 (31%), Positives = 68/152 (44%)

Query: 364 APSWPSLVDTSKQWDYYARREKDRDR-ERDRDRERDRDRDRERERTREERERERDHSPTPS 422
AP P + T + + E RD R+ + RD + E E+ + +E+ER
Sbjct: 143 APLIPYPLITKEDINAIEEEDKRDLSREISKFRDTHKKLEEEKGK-KEKERQEIEKER 201

Query: 423 VFNSEERYRYREYAERGYERHRA-SREKE-ERHRER-RHREKEETHKSSRSNSRRRH 478
+ ER R RE ER ER R REKE ER RER R R+++ T+ + R R R
Sbjct: 202 R-ERDRDRDRDRDRDRDRDRDRDRDR 260

Query: 479 ESEEGDSHRRHKHKKSRSKEGKEAGSEPAPEQE 512
E S R +S+ +E E E+E
Sbjct: 261 RDRERSSDRNKDRSRSEKSRDRERERERERE 294

Score = 126 (18.9 bits), Expect = 1.8e-04, Sum P(2) = 1.8e-04
Identities = 41/149 (27%), Positives = 66/149 (44%)

Query: 375 KQWDYYARREKDRDRDRDRDRERDRDRERERTREERERERDHSPT---PSVFNSE--EE 429
K W+ R+K R+ E++ +RE +R R+ +E R +E D+ P + ++
Sbjct: 354 KNWEI-RERKKTREYEKEAEREERREMAKEAKRLKEFLEDYDDDRDDPKYYRGSAQK 412

Query: 430 RYRYREYAERGYERHRASREKEERHRER-----HREKEETHKSSRSNSRRRHSE--E 481
R R RE ER R REKEE R+ H + + + + RRR +
Sbjct: 413 RLRDREKEMEADERDR-KREKEELEIRQLLAEGHPDPDAELQRMQEAEARRRQPQIKQ 471

Query: 482 EGDHSHRRHKHKKSRSKEGKEAGSEPAPEQE 512
E +S + K+ K K + E PEQ+
Sbjct: 472 EPESEEEEEEKQEKEEKREPEMEEEEEPEQK 502

Score = 124 (18.6 bits), Expect = 3.0e-04, Sum P(2) = 3.0e-04
Identities = 41/141 (29%), Positives = 65/141 (46%)

Query: 380 YARREKDRD-RERDRDRERDRDRDRERERTREERERERDHSPTPSVFNSEERYRYREYAE 438
Y R K+ + RER + RE +++ +RE ER RE +E + + D++R + Y
Sbjct: 349 YQERLKNWEIRERKKTREYEKEAEREERREMAKEAKRLKE-FLEDYDDDRDDPKYYRG 407

Query: 439 RGYERHRASREKEERHRER-RHREKEETHKSSRSNSRRRHSEEGDSHRRHKHKKSRS 497
++ REKE ER R REKEE R + H + + R + + +R
Sbjct: 408 SALQKRLRDREKEMEADERDRKREKEELEIRQLLAEG-HPDPAELQRMQEAEARRRQ 466

Query: 498 KEGKEAGSEPAPEQESTATPAE 520
+ K+ EP E+E E E
Sbjct: 467 PQIKQ--EPESEEEEEEKQEKE 486

Score = 121 (18.2 bits), Expect = 6.2e-04, Sum P(2) = 6.2e-04
Identities = 43/149 (28%), Positives = 67/149 (44%)

Query: 364 APSWPSLVDTSKQWDYYARREKDRDR-ERDRDRERDRDRDRERERTREERERERDHSPTPS 422
AP P + T + + E RD R+ + RD + E E+ + +E+ER
Sbjct: 143 APLIPYPLITKEDINAIEEEDKRDLSREISKFRDTHKKLEEEKGK-KEKERQEIEKE- 200

Query: 423 VFNSEERYRYREYAERGYERHRASREKEERHRERHREKEETHKSSRSNSRRRHSEEE 482
+ ER R RE R ER R RE+E + R RE+E R + R+ R R E
Sbjct: 201 --RRERERERERERERERERERERERERERERERERERERERDRDRD-RTKERDRDRDRE 256

Query: 483 GDSHRRHKHKKSRSKEGKEAGSEPAPEQE 512
D R + + S R+K+ + E + ++E
Sbjct: 257 RDRDR-DRERSSDRNKD-RSRSEKSRDR 284

Score = 105 (15.8 bits), Expect = 3.1e-02, Sum P(2) = 3.1e-02

```

Identities = 25/73 (34%), Positives = 33/73 (45%)

Query:   428 EERYRYREYAERGYERHRASREKE-ERHRERRHREKEETRHKSSRSNSRRRHSEEGDSH 486
      EE   +E   + E+ R RE+E ER RERR RE+E R +       R E E
Sbjct:   184 EEEKGKKEKERQEIEKERRERERERERERERREREREREREKEKERERERERDRDR 243

Query:   487 RRHKHKKSKRSKE 499
      R K +   R +E
Sbjct:   244 DRTKERDRDRDRE 256

Score = 105 (15.8 bits), Expect = 3.1e-02, Sum P(2) = 3.1e-02
Identities = 31/87 (35%), Positives = 45/87 (51%)

Query:   382 RREKDRDRERDRDRERDRDRDRER-ERTREERERERDHSPTSPVFNSEERYRYREYAERG 440
      +R +DR++E + D ERDR R++E E R+R   H P P       D E R + AER
Sbjct:   412 KRLRDREKMEAD-ERDRKREKEELEETIRQLLAEGH-PDP----DAELQRMQEQAERR 464

Query:   441 YERHRASREKEERHRERRHREKEETRHK 468
      + +   +E E   E +EKEE R +
Sbjct:   465 -RQPQIKQEPESEEEEEEKQEKEEKREE 491

Score = 46 (6.9 bits), Expect = 1.5e-16, Sum P(2) = 1.5e-16
Identities = 13/49 (26%), Positives = 21/49 (42%)

Query:   54 AENGVPKPKVTETEDDSDSDSDDDDDVHVHTIGDIKTGAPQYGSYGTAP 102
      A NG +P+   +D+   D +   D + G I+   +Y S   AP
Sbjct:   70 ASNGNARPETVTNDDEEALDEETKRRDQMIK-GAIEVLIREYSSSELNAP 117

Score = 46 (6.9 bits), Expect = 1.8e-04, Sum P(2) = 1.8e-04
Identities = 14/53 (26%), Positives = 21/53 (39%)

Query:   30 ENEVERPEEENASANPPSGIEDETAENGVPKPKVTETEDDSDSDSDDDDDVHV 82
      + E ER E   E E E   + + E E D   D ++DE+D +
Sbjct:   282 DRERERERERERERERERERERERERER-EREREREREREREKDKKRDREDEEDAY 333

Score = 44 (6.6 bits), Expect = 2.0e-13, Sum P(2) = 2.0e-13
Identities = 13/60 (21%), Positives = 21/60 (35%)

Query:   20 DEEEEWLYGDENEVERPEEENASANPPSGIEDETAENGVPKPKVTETEDDSDSDSDDD 79
      ++E + +   + E ER E   + E   K + E E D D D + D
Sbjct:   191 EKEROEIKERREREREREREREREREREREREREREKEKERERERERDRDRDRTKRD 256

```

Pedant information for DKFZphutel_17k7, frame 3

Report for DKFZphut1 17k7.3

```
[LENGTH]          520
[MW]               58375.30
[pI]              5.41
[HOMOL]           PIR:S62454 hypothetical protein SPAC22G7.10 - fission yeast
                  (Schizosaccharomyces pombe) 3e-18
[FUNCAT]          04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) (S.
cerevisiae, YJR093c) 2e-13
[FUNCAT]          30.10 nuclear organization (S. cerevisiae, YJR093c) 2e-13
[PROSITE]         MYRISTYL 9
[PROSITE]         AMIDATION 1
[PROSITE]         CK2_PHOSPHO_SITE 18
[PROSITE]         TYR_PHOSPHO_SITE 2
[PROSITE]         PKC_PHOSPHO_SITE 12
[PROSITE]         ASN_GLYCOSYLATION 2
[KW]              Alpha_Beta
[KW]              LOW COMPLEXITY 35.00 %
```

```

SEQ      MSAGEVERLVSELSGGTGDEEEEWLYGDENEVERPEEENASANPPSGIEDETAENGVPK
SEG      .....XXXXXXXXXXXX.....
PRD      cccchhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      PKVTETEDSDSDSDDDDDDVHVTIGDIKTGAPQYGSYGTAPVNLNIKTGGRVYGTGTGK
SEG      ..XXXXXXXXXXXXXXXXXXXX.....
PRD      cceeeccccccccccccccccceeeccccccccccccccccccccceeecccccecccccc

SEQ      VKGVDLDAPGSINGVPLLEVDLDSFEDKPRWPGADLSDYFNYGFNEDTWKAYCEKQKRI
SEG      .....
PRD      ceeeccccccccccccceeeccccccccccccccccccccccccccccccccchhhhhhhhhhhh

SEQ      RMGLEVIPVTSTTNKITVQQRTGNSEKETALPSTKAEFTSPPSLFKTGLPPSRRLPGAI
SEG      .....

```

```
PRD      hhhheeeeeccccceeeeeeeccccccccccccccccceeeccccceeecccccccccccc
SEQ      DVIGQTITISRVGRRRANENSNIQVLSERSATEVDNFSKPPFFFPAGPPTHLPFFPF
SEG      .....XXXXXXXXXXXXXXXXXXXXX
PRD      cccccceeeeeccccccccccccceeecccccccccccccccccccccccccccccccccc

SEQ      LPPPTVSTAPPLIPPPGFPFPFGAPPSLIPTIESGHSSGYDSRSARAFYPYGNVAFPHL
SEG      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccceeeccc

SEQ      PGSAPSWSPLVDTSKQWDYYARREKDRDRERDRDRERDRERERERERERDHSPT
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccceeeccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhhhcccccc

SEQ      PSVFNSEDEERYRYREYAERGYERHRASREKEEHRERRHREKEETRHKSSRSNSRRRHES
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccchhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccc

SEQ      EEGDSHRRHKKHKKSKRSKEGKEAGSEPAPEQESTEATPAE
SEG      XX..XXXXXXXXXXXXXXXXX.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphutel 17k7.3

PS000001	40->44	ASN_GLYCOSYLATION	PDOC000001
PS000001	278->282	ASN_GLYCOSYLATION	PDOC000001
PS000005	169->172	PKC_PHOSPHO_SITE	PDOC000005
PS000005	193->196	PKC_PHOSPHO_SITE	PDOC000005
PS000005	206->209	PKC_PHOSPHO_SITE	PDOC000005
PS000005	214->217	PKC_PHOSPHO_SITE	PDOC000005
PS000005	233->236	PKC_PHOSPHO_SITE	PDOC000005
PS000005	268->271	PKC_PHOSPHO_SITE	PDOC000005
PS000005	346->349	PKC_PHOSPHO_SITE	PDOC000005
PS000005	373->376	PKC_PHOSPHO_SITE	PDOC000005
PS000005	469->472	PKC_PHOSPHO_SITE	PDOC000005
PS000005	474->477	PKC_PHOSPHO_SITE	PDOC000005
PS000005	485->488	PKC_PHOSPHO_SITE	PDOC000005
PS000005	494->497	PKC_PHOSPHO_SITE	PDOC000005
PS000006	2->6	CK2_PHOSPHO_SITE	PDOC000006
PS000006	17->21	CK2_PHOSPHO_SITE	PDOC000006
PS000006	47->51	CK2_PHOSPHO_SITE	PDOC000006
PS000006	64->68	CK2_PHOSPHO_SITE	PDOC000006
PS000006	66->70	CK2_PHOSPHO_SITE	PDOC000006
PS000006	70->74	CK2_PHOSPHO_SITE	PDOC000006
PS000006	72->76	CK2_PHOSPHO_SITE	PDOC000006
PS000006	74->78	CK2_PHOSPHO_SITE	PDOC000006
PS000006	84->88	CK2_PHOSPHO_SITE	PDOC000006
PS000006	144->148	CK2_PHOSPHO_SITE	PDOC000006
PS000006	206->210	CK2_PHOSPHO_SITE	PDOC000006
PS000006	215->219	CK2_PHOSPHO_SITE	PDOC000006
PS000006	250->254	CK2_PHOSPHO_SITE	PDOC000006
PS000006	271->275	CK2_PHOSPHO_SITE	PDOC000006
PS000006	273->277	CK2_PHOSPHO_SITE	PDOC000006
PS000006	340->344	CK2_PHOSPHO_SITE	PDOC000006
PS000006	369->373	CK2_PHOSPHO_SITE	PDOC000006
PS000006	426->430	CK2_PHOSPHO_SITE	PDOC000006
PS000007	434->442	TYR_PHOSPHO_SITE	PDOC000007
PS000007	152->161	TYR_PHOSPHO_SITE	PDOC000007
PS000008	15->21	MYRISTYL	PDOC000008
PS000008	96->102	MYRISTYL	PDOC000008
PS000008	115->121	MYRISTYL	PDOC000008
PS000008	130->136	MYRISTYL	PDOC000008
PS000008	154->160	MYRISTYL	PDOC000008
PS000008	229->235	MYRISTYL	PDOC000008
PS000008	244->250	MYRISTYL	PDOC000008
PS000008	289->295	MYRISTYL	PDOC000008
PS000008	362->368	MYRISTYL	PDOC000008
PS000009	253->257	AMIDATION	PDOC000009

(No Pfam data available for DKFZphut1 17k7.3)

DKFZphut1_18c12

group: uterus derived

DKFZphut1_18c12 encodes a novel 378 amino acid protein nearly identical to human WUGSC:H_DJ0872F07.1 protein.

The novel protein has an additional N-terminal domain, which is not present in WUGSC:H_DJ0872F07.1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

nearly identical to human WUGSC:H_DJ0872F07.1 protein

on genomic level encoded by AC004537, 10 exons the predicted protein sequence AC004537_1 is only partially o.k. first exon wasn't predicted there are additional exons predicted (BLASTX/EST-BLAST shows that the cDNA is only partly spliced) intron ~1216-3540//~3577-5059

Sequenced by AGOWA

Locus: map="7q31"

Insert length: 6005 bp

Poly A stretch at pos. 5980, polyadenylation signal at pos. 5968

```

1 AGCGGGTGCT GCTAGCGGAG GCGCCATATT GGAGGGGACA AAACCTCCGGC
51 GACAGCGAGT GACACAAATA AACCCCTGGA CCCCTTGTT CCCTCAGCTC
101 TAAGGGCCGC GATGTTGTAC CTAGAAGACT ATCTGGAAAT GATTGAGCAG
151 CTTCCATATG ATCTGCGGGA CCGCTTCACG GAAATGCGCG AGATGGACCT
201 GCAGGTGCAG AATGCAATGG ATCAACTAGA ACAAAGAGTC AGTGAATTCT
251 TTATGAATGC AAAGAAAAAT AAACCTGAGT GGAGGGAAGA GCAATGGCA
301 TCCATCAAAA AAGACTACTA TAAAGCTTTG GAAGATGCAG ATGAGAAGGT
351 TCAGTTGGCA AACCAGATAT ATGACTTGGT AGATCGACAC TTGAGAAAGC
401 TGGATCAGGA ACTGGCTAAG TTTAAATGG AGCTGGAAGC TGATAATGCT
451 GGAATTACAG AAATATTAGA GAGGCGATCT TTGGAATTAG ACACCTCCTC
501 ACAGCCAGTG AACAAATCACC ATGCTCATTC ACATACTCCA GTGGAAAAAA
551 GGAATATATA TCCAACCTCT CACCATACGA CAACAGATCA TATTCCTGAA
601 AAGAAATTTA AATCTGAAGC TCTTCTATCC ACCCTTACGT CAGATGCCTC
651 TAAGGAAAAA ACACTAGGTT GTCGAAATAA TAATTCCACA GCCTCTTCTA
701 ACAATGCCTA CAATGTGAAT TCCTCCCAAC CTCTGGGATC CTATAACATT
751 GGCCTCGTTAT CTTCAGGAAC TGGTGCAGGG GCAATTACCA TGGCAGCTGC
801 TCAAGCAGTT CAGGCTACAG CTCAGATGAA GGAGGGACGA AGAACATCAA
851 GTTTAAAGAC CAGTTATGAA GCATTTAAGA ATAATGACTT TCAGTTGGGA
901 AAGAATTTTT CAATGGCCAG GGAACAGTT GGCTATTCAT CATCTTCGGC
951 ACTTATGACA ACATTAACAC AGAATGCCAG TTCATCAGCA GCCGACTCAC
1001 GGAGTGGTCG AAGAGGAAAA AACAAACAAC AGTCTTCAAG CCAGCAGTCA
1051 TCATCTTCCT CCTCCTCTTC TTCCTTATCA TCGTGTTCCT CATCATCAAC
1101 TGTGTACAAA GAAATCTCTC AACAAACAAC TGTAGTGCCA GAATCTGATT
1151 CAAATAGTCA GGTGATTGG ACTTACGACC CAAATGAACC TCGATACTGC
1201 ATTTGTGAAT AGGTAAAGT CTGTTATATC TATAAAAGTA TAATCTGAAT
1251 AACTAGAAG GAAGAGAACT ATTTCAATTT TAAGCACTTT TTTAAACTCA
1301 CTTAAATATC CTTTGCCTTA TTTGTATACT TTTCTCCCCC TTCTTACAAA
1351 AGTGACATTT GCTGTAATA CTGAGTATAA AGAAAAATGT TACCCATAAT
1401 CTAGCCCTCT AGATACAACC TGTAACATAA CATTTTGGT ATACCACTAC
1451 CATATACCTC ATGTGCACAT TGGCTGCCTT AATAAAATAC AACAGACTGG
1501 GTAGCTTAAA CAACAGAAAA TAATTTTCTC ACAGGTATGA AGGCTGGGAA
1551 GTCCAAGATC AAGGTGTCCA CTGACTCAGT TCTGGAGGAG GGCTCCCTTC
1601 CTAGATGGAG ACTGCTGCCT TCTCACCAGG TCCTCACATG ATAGAGGGAG
1651 AAAGAGTGTG CTCTGGTGTG TTTTCTTATA AGGGCACCAG CCTTGTGAGA
1701 GTAGGACCCC ACTCTATGAC CTCATTTAAC CTTTACCACC TCCTCACAGG
1751 CCTGTGTTCC AATTATAGTC ACGTGGGGGG TTAGGGCTTC AACATATGAT
1801 TTTGAGACAT AAGCTTGCAT TTCATAACAC GTGTCTATGC AGATTTGCAC
1851 ATGCATGTGT GTATAAGTTT GTCAGTAGGA ACCACAGTGT ATACTTTCTT
1901 GTTATGAAAT TTTTCTCTA AATCAGGTAT ACCGAACATG ATTTTCTTTT
1951 AAGATCATAT TTTAATTTT CACATAGTTA TCTCTTATGC CATCCAGTGT
2001 AGTTTCTTTA ACCAATACCT AGCTATAGAT TATATTAGTG GTTTTAAATT
2051 GTTTGAAATT AGGGATAATA TTACGATAGG CATTTTAAAT ATGTAATCCA
2101 TTTTATACAT CTAATTTCTT GGATAATCTT TTAGAAATAA AATTAGGCTG
2151 TAAATATTTT ACAGACACCA AAATATATTT TCTAGAAATT TATTACCAAA
2201 AATTAATAAA CATACCGGTT TACTAAACCC TGTCACACAC TGGATATTAT
2251 TTTCTTTTAA AACTAAGTA CCAATTTGGT AGTTTATATAT TATGATTGTT
2301 TTAATACAC TAGTATTATT GAAGTTGGAC ATTTTGTGAC CATTTTGTGT
2351 TTTTACATTA TGAATCGACT CCTAATGGTG TCGGCTGATT TTTCTATTGT

```

```

2401 TTTTGTATG TACTCTAAAT ATTTGCTTGA TTTAGTTTTT TAAAAATAAT
2451 TCTAAAAATTT TAATTTTATG TAGTTATGAC TGTTAAATTT TTTTATGAA
2501 GCAAGCCATG GATTATATAC TTAGAAGGGC TTTCTCTTTG GCTCTTCTTT
2551 CTACAAAAAA TTGCTTTGTA TAATATTTTC TCCTAGTTTT TATATGGTTT
2601 TGCTAGTTTC TTTGCATGCT TCAGTTTCTT CACATTTAAG ACTTAGTCTA
2651 TCAGCAGATT ATTTGTGTCTA ACAGTATGAG TTGCCAGTCT GATTTTTAAA
2701 AATTTTAACA ATTTGTTAGC TGTTCCTACT TCACCCGATA AACATTTTTC
2751 AGTACAAATG ATAGAAAAGC ATATCCTGTA TCCTGACAAC AAAAGTAGAT
2801 TACTTGCAAA AGAACAAAAT CAGACTGAAC CTAGAGTTTT CCTCTGTAAC
2851 ACTAAAAAAC TAGAAGGTGA TGGAAATATG CTGTAGAGCT TTCAGGGAAA
2901 AATTAAGAGC CCCCAAAAAC TTGATATTCA GAGAAGTTAT TTCTCTGCAT
2951 AGGACCATTG AAAATATATT TCACTCATGC AGAGAATCAG AAGATATGCC
3001 ATCTAGTTAA TCCTGTCTGA AAAATTATTC AATCCACTGA GAACTTTCAGT
3051 GAACTCAAGA ATTAGCAAGT TATGCCCTAA AGTGCTGGTG ATGAAGAGCA
3101 AAAGAAAAAT GAGAAAGGAC ATAAAAATGA TAAGTTTAGA AGTTTCAAGG
3151 AAGGAGACTA TTAATTGCAA AAATATATAT GACCTAATGT GACCCAAGAA
3201 GTAAAAAACC TCAGTAAGTA AATAATCAAG AAAGGAACCT AAAATTTTAA
3251 CAATAAGAAC TACCAGAAA GATGACTCCT TCATCCGGGT GATTTATATG
3301 TCAAGTTCTT CCAGACTTCT GAAGGGCAGA TAATTCCTGT GCATTTCCTC
3351 CCACCCTTGC CCCACCCTGC CCAAAAGAGT ATTTGAGAA AAAATTATTA
3401 TACCTTGATT CTCAAATGTA TTGTATATTC AGTGTATTTT CCTTTATTTT
3451 CAGGCAAGTAT CATACATAAA CAGTTAATTG GTATCTAGGT GTTTGTTACA
3501 TAGTCATAAT AAAGACATTT AATTTTTTTT AACTAGGTAT CTTATGGTGA
3551 GATGGTGGGA TGTGATAACC AAGATGTAAG TATTACATTT TTCTATTTAG
3601 GAATGAAAAA AATCACAGGT TGTATTACT TGAATATTG TCTTATTGTC
3651 TGTATGGTTT GGTCTAAGAA AACAGGTTTG CAGGTATATT AGTTATGTTA
3701 TGCCTAATGCT AGAATATTCC TCTTCAAAAT AGGGTAGTGT CCTTAATGT
3751 GTTCCCTATT TTAATTTTTA AAGCTAATTT TATGGTTTTA TGTGCAGATT
3801 GTCTCAGAAG TGTATGTGTG TATGAAAATT ATAAATACCC TCCTTCCCTT
3851 TTACTAAAAA ATACTGTGTT TACTAGAATC CAGTTCATTT ATCACAATGA
3901 AGAAATGGAA TTTTAAAACA ATTCATTCTT TCAGGCTGCA CCGTGCTAAA
3951 GTGAAGGGTG GGATAATGTA GGATCTAATG TGAGATTATC TTCCTCTCAT
4001 GAGTATAATA TTTTTCCTG TACTCTGCAG GTGTCAGCTG ATAAGAGCCA
4051 CCCCTGATCT AAAAAGTAAA GGAAATTTGA AAGGAAGGAA TTCTTGTTTT
4101 TTAGGAGACT TAATTTTAGT TAGAGATACG TTTTATTATC AATACTGAGA
4151 ATATTTGTTG CTAGTAATTT TGACTCCCTC CTTATTTAGT AGTGACAGGA
4201 TCCTAAGATT AACAAAGATT TTAATTTGT AAAACAATCT GAAGATTGAG
4251 GGAGCTGGCT AGGTGCATTA AAATGTGTAC TTTTCTTAGA CCTGATAGGG
4301 TTACAGCAAC ATGCTCACGT AGATTGGGAC AGAGCCTCCT TCTGTTTCCC
4351 TGCTAGAAAT CCCTGTAGG CTGTTTGTGG TTGTTGCAAA AACAAATATTG
4401 CCCAACCAT TCAAGAACAT CACTGTAAAC TCTTCTGGGG CAGTTAGTGA
4451 AAATGATGAA TGAGATTCTT ATGAGTACCA GCATCATGCT TCTCTGATTC
4501 TTCTTATTC CAGTTGTGCT CTTCTGAGTG CTAAGACTTT CATGAAAGAG
4551 TTTTCTGCTT AATATGTTT AAAGAGGAAT AATTTTCTC TACATTTCAA
4601 GGAATAGAAA CACCCACGTA GGAAATGCAG GGCATAAGAC ATAAATTAAT
4651 GTCTTTAATT ACAATCAGCT TATTCTACTT TATGAGACAG CAAATAAGGC
4701 TGACTATTAA ATAAATCTT AAGTTATATT TACCTTCTAC ATAGAAGATT
4751 CATCCACTT CTTTGTGCCC TTGAAAGCTG AAAACTAGTG AATTTTCATT
4801 CATTAGGATG AGGGGACTAG ATTACATGGA CCTCAGGATT CTTGAAGATG
4851 CATAAATTTT CTGTGCCTTC ATTTCCCTCAT TCCTGAAGCT TATCATTTAG
4901 TCTAAATGAT GTCTAAATAA TCTAGATCTA AAAATTCTGA TGTACACAT
4951 CTAATTATTG TTAATTTAAA TGGATTATTC AGTCCTCTGA GCATATTTTA
5001 ATATACTCTC TTGTCTTCAG AAGTACTGAA AACTTGTTTT TTGCAATTTT
5051 GCTTCTTAGT GCCCTATAGA ATGGTTCAT TATGGCTGCG TTGGATTGAC
5101 AGAGGCACCA AAAGGCAAA GGTACTGTCC ACAGTGCACT GCTGCAATGA
5151 AGAGAAGAGG CAGCAGACAC AAATAAAGGT GGTCTTTTGT TTTGATGAAG
5201 AAATAAATCT CAGCTGAAGA TTTTATATAG GACTTTAAAA AGAAGAGAAG
5251 AGAAAGAAGA AACAAATGCAT TTCCAGGCAA CCACTTAAAG GATTTACATA
5301 GACAATCTTA TAAGATCTTG AACTTGAATT TTATGGGTTG TATTTAATA
5351 ATGTAAGTAA ATTATTTATG CACTCCTGGT GTGCTATGAA TATTATTCCA
5401 GTTAGCCTTG GATTATTTCA GTGGCCAACA TATGCAGACA TTTGTACTCC
5451 TCAACCATTT TCTCAAAGTA ATGGGCATTC TATGATTAG ACTTCAAGGA
5501 ATTCCAATGA TGAAGATTTT AAGGAAAGTA TTTTATATTC AACAGGTATA
5551 TTCTGCTGCA TGTACTGTAC TCCAGAGCTG TTATGTAACA CTGTATATAA
5601 ATGGTTGCAA AAAAAAATAA AAGTCAGTGC TTCTAAAAG AATTAAAGAT
5651 AATGGTTTTT AAAATGCCCT TATAATAAGC TTTGTTTCTT TGTGAACTA
5701 ATTCAGCAGG CTGAAGGAAA TGGTTCATGT GATAATGTGG GCTGGTATCC
5751 TCTAGAGTAC CTGGGTACAT AAACAGAAAC TCCTGTAGGT AAAAAGTAAT
5801 TTGTGCCATT AGTCTTTCTA TGTCTTCTGA TCCAGATAGA GTGCAGTTCA
5851 TTGAGGAGGG GGCGGGGGAC TGAAGGGGAA AGGGCGTTAA AGTGATACAT
5901 TTTTATACCA AATGTGTTTA TTTTTTTGTG CAAGTAATCC TTAATTTGTC
5951 AATTGTATTA GGTGTTAAAA TAAAGTTTTT AAAAAATTA AAAAAAATAA
6001 AAAAA

```

BLAST Results

Entry HSG20547 from database EMBL:
HSG20547| human STS A005W09.
Length = 154

Minus Strand HSPs:
 Score = 770 (115.5 bits), Expect = 2.9e-26, P = 2.9e-26
 Identities = 154/154 (100%)

Medline entries

98101645:
 The candidate tumour suppressor p33ING1 cooperates with p53 in cell growth control.

Peptide information for frame 1

ORF from 112 bp to 1245 bp; peptide length: 378
 Category: similarity to known protein

```

1 MLYLEDYLEM IEQLPMDLRD RTEMREMDL QVQNAMDQLE QRVSEFFMNA
51 KKNKPEWREE QMASIKKDY KALEDADEKV QLANQIYDLV DRHLRKLQDE
101 LAKFKMELEA DNAGITEILE RRSLELDTPS QPVNNHHAHS HTPVEKRKYN
151 PTHSHHTTDH IPEKKFKSEA LLSTLTSDAS KENTLGCRNN NSTASSNNAY
201 NVNSSQPLGS YNIGSLSSGT GAGAITMAAA QAVQATAQMK EGRRTSSLKA
251 SYEAFKNDF QLGKEFSMAR ETVGYSSSSA LMTTLTQNAS SSAADSRSGR
301 KSKNNKSSS QSSSSSSSS SLSSCSSST VVQEISQQT VVPESDSNSQ
351 VDWTYDPNEP RYCICNQKV CYIYKSII

```

BLASTP hits

Entry AF044076.1 from database TREMBL:
 "ING1"; product: "candidate tumor suppressor p33ING1"; Homo
 sapiens candidate tumor suppressor p33ING1 (ING1) mRNA, complete
 cds. Homo sapiens (human)
 Length = 279
 Score = 162 (57.0 bits), Expect = 1.1e-09, P = 1.1e-09
 Identities = 48/183 (26%), Positives = 92/183 (50%)

Entry AC004537.1 from database TREMBL:
 gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07 from
 7q31, complete sequence.
 Score = 1814, P = 3.7e-187, identities = 358/358, positives = 358/358

Entry CEY51H1A.1 from database TREMBL:
 gene: "Y51H1A.4"; Caenorhabditis elegans cosmid Y51H1A
 Score = 213, P = 3.7e-15, identities = 37/123, positives = 82/123

Alert BLASTP hits for DKFZphut1_18c12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_18c12, frame 1

Report for DKFZphut1_18c12.1

```

[LENGTH]      378
[MW]           42275.72
[pI]           5.72
[HOMOL]        TREMBL:AC004537.1 gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07
from 7q31, complete sequence. 1e-157
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YHR090c] 8e-05
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YNL097c] 2e-04
[PROSITE]      MYRISTYL 3
[PROSITE]      AMIDATION 2
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 4
[PROSITE]      PROKAR_LIPOPROTEIN 1
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 3
[PROSITE]      ASN_GLYCOSYLATION 5
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 20.63 %

```

```
[KW] COILED_COIL 7.94 %
```

```
SEQ MLYLEDYLEMIEQLPMDLRDRFTEMREMLDQVQNAMDQLEQRVSSEFFMNAKKNKPEWREE  
SEG .....  
PRD ccchhhhhhccccccchhhhhhccccccccccccccccccccccccccccccccccccccc  
COILS .....
```

```
SEQ QMASIKKDYYKALEDADEKVQLANQIYDLVRHLRLKLDQELAKFKMELEADNAGITEILE  
SEG .....  
PRD hhhhhhhhccccccchhhhhhccccccccccccccccccccccccccccccccccccccc  
COILS .....CCCCCCCCCCCCCCCCCCECCCECCCEC.....
```

```
SEQ RRSLELDTSPQPVNHHASHPTVEKRKYNPSTSHHTTTDHIPEKKFKSEALLSTLTSDAS  
SEG .....  
PRD hhccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc  
COILS .....
```

```
SEQ KENTLGCRNNNSTASSNNAYNVNSSQPLGSYNIGSLSSGTGAGAITMAAAQAVQATAQMKB  
SEG ..... xxxxxxxxxxxxxxxxxx ..... xxxxxxxxxxxxxxxxxxxxxxxx..  
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc  
COILS .....
```

```
SEQ EGRRTSSLKASYEAFKNPDFQLGKEFSMARETVGYSSSSALMTTLTQNASSSAADSRSRGR  
SEG ..... xxxxxxxxxxxxxxxx  
PRD hccccccccchhhhhhccccccccccccccccccccccccceeeecccccccccccccccc  
COILS .....
```

```
SEQ KSNNNKSSSQSSSSSSSSLSSCSSSVVQEISQQTTVPESDSNSQVDWTYPNEP  
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxx  
PRD cccccccccccccccccccccceecccccccccccccccccccccccccceeeccccc  
COILS .....
```

```
SEQ RYCICNQVKVCIIYSII  
SEG .....  
PRD eeeeeeeeeeeeccc  
COILS .....
```

Prosites for DKFZphut1 18c12.1

PS00001	190->194	ASN_GLYCOSYLATION	PDOC00001
PS00001	191->195	ASN_GLYCOSYLATION	PDOC00001
PS00001	203->207	ASN_GLYCOSYLATION	PDOC00001
PS00001	288->292	ASN_GLYCOSYLATION	PDOC00001
PS00001	306->310	ASN_GLYCOSYLATION	PDOC00001
PS00002	218->222	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	243->247	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	64->67	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	298->301	PKC_PHOSPHO_SITE	PDOC00005
PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	349->353	CK2_PHOSPHO_SITE	PDOC00006
PS00008	186->192	MYRISTYL	PDOC00008
PS00008	214->220	MYRISTYL	PDOC00008
PS00008	219->225	MYRISTYL	PDOC00008
PS00009	241->245	AMIDATION	PDOC00009
PS00009	298->302	AMIDATION	PDOC00009
PS00013	315->326	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphut1 18c12.1)

DKFZphutel_18i19

group: transcription factors

DKFZphutel_18i19 encodes a novel 759 amino acid protein with similarity to the SREBP-2 mutant sterol regulatory element binding protein-2 of *Cricetulus griseus*.

The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH₂-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

The new protein can find application in modulating/blocking the expression of genes involved in lipid metabolism.

similarity to transcription factor SF3

complete cDNA, complete cds, EST hits
strong similarity to mutated SREBP-2 of hamster,
similarity is not to SREP-2 part of protein but to the unknown part of
the fusion protein

Sequenced by AGOWA

Locus: /map=12

Insert length: 3664 bp

Poly A stretch at pos. 3647, polyadenylation signal at pos. 3636

```

1 GCGCTAGGTA GAGCGCCGGG ACCTGTGACA GGGCTGGTAG CAGCGCAGAG
51 GAAAGGGCGG TTTTAGCCAG GTATTTCAGT GTCTGTAGAC AAGATGGAAT
101 CATCTCCATT TAATAGACGG CAATGGACCT CACTATCATT GAGGGTAACA
151 GCCAAAGAAC TTTCTCTTGT CAACAAGAAC AAGTCATCGG CTATTGTGGA
201 AATATTCTCC AAGTACCAGA AAGCAGCTGA AGAAACAAAC ATGGAGAAGA
251 AGAGAAGTAA CACCGAAAAT CTCTCCAGC ACTTAGAAA GGGGACCCCTG
301 ACTGTGTAA AGAAGAAGTG GGAGAACCCA GGGCTGGGAG CAGAGTCTCA
351 CACAGACTCT CTACGGAACA GCAGCACTGA GATTAGGCAC AGAGCAGACC
401 ATCCTCTCTG TGAAGTGACA AGCCACGCTG CTTCTGGAGC CAAAGCTGAC
451 CAAGAAGAAC AAATCCACCC CAGATCTAGA CTCAGGTCAC CTCCTGAAGC
501 CCTCGTTCA GGTGATATC CCCACATCAA GGACGGTGAG GATCTTAAAG
551 ACCACTCAAC AGAAAGTAAA AAAATGGAAA ATTGTCTAGG AGAATCCAGG
601 CATGAAGTAG AAAAATCAGA AATCAGTGAA AACACAGATG CTTCCGGCAA
651 AATAGAGAAA TATAATGTTT CGCTGAACAG GCTTAAGATG ATGTTTGAGA
701 AAGGTGAACC AACTCAAAC AAGATTCTCC GGGCCCAAAG CCGAAGTGCA
751 AGTGGAAAGG AGATCTCTGA AAACAGCTAT TCTCTAGATG ACCTGGAAAT
801 AGGCCCAGGT CAGTTGTCAT CTTCTACATT TGACTCGGAG AAAAATGAGA
851 GTAGACGAAA TCTGGAACCT CCACGCCTCT CAGAAACCTC TATAAAGGAT
901 CGAATGGCCA AGTACCAGGC AGCTGTGTCC AAACAAAGCA GCTCAACCAA
951 CTATACAAAT GAGCTGAAAG CCAGTGGTGG CGAAATCAAA ATTCAATAAA
1001 TGGAGCAAAA GGAGAATGTG CCCCAGGTC CTGAGGTCAG CATCACCCAT
1051 CAGGAAGGGG AAAAGATTTC TGCAAAATGAG AATAGCCTGG CAGTCCGTTT
1101 CACCCCTGCC GAAGATGACT CCGTGACTC CCAGGTAAAG AGTGAGGTTT
1151 AACAGCCTGT CCATCCCAAG CCACTAAGTC CAGATCCAG AGCCTCCAGT
1201 CTTTCTGAAA GTTCTCCTCC CAAAGCAATG AAGAAGTTTC AGGCACCTGC
1251 AAGAGAGACC TGCGTGGAAAT GTCAGAAGAC AGTCTATCCA ATGGAGCGTC
1301 TCTTGGCCAA CCAGCAGGTG TTTACATCA GCTGCTCCG TTGCTCCTAT
1351 TGCAACAACA AACTCAGTCT AGGAACATAT GCATCTTAC ATGGAAGAAT
1401 CTATTGTAAG CCTCACTTCA ATCAACTCTT TAAATCTAAG GGCAACTATG
1451 ATGAAGGCTT TGGGCACAGA CCACACAAGG ATCTATGGGC AAGCAAAAAT
1501 GAAAACGAAG AGATTTTGA GAGACCAAGC CAGCTTGCAA ATGCAAGGGA
1551 GACCCCTCAC AGCCCAGGGG TAGAAGATGC CCTATTGCT AAGGTGGGTG
1601 TCCTGGCTGC AAGTATGGAA GCCAAGGCCT CCTCTAGCA GGAGAAGGAA
1651 GACAAGCCAG CTGAAACCAA GAAGCTGAGG ATCGCCTGGC CACCCCCAC
1701 TGAAGTGAAG AGTTCAGGAA GTGCCCTGGA GGAAGGGATC AAAATGTCAA
1751 AGCCCAAAAT GCCTCTGAA GACGAAATCA GCAAGCCGA AGTTCCTGAG
1801 CATGTGCGATC TAGATCTGAA GAAGCTAAGA CGATCTTCTT CACTGAAGGA
1851 AAGAAGCCGC CCATTCACTG TAGCAGCTTC ATTTCAAAGC ACCTCTGTCA
1901 AGAGCCCAAA AACTGTGTCC CCACCTATCA GGAAGGGCTG GAGCATGTCA
1951 GAGCAGAGTG AAGAGTCTGT GGGTGGAGA GTTGCAGAAA GGAACAAGT
2001 GGAAGATGCC AAGGCTTCTA AGAAGAATGG GAATGTGGGA AAAACAACCT
2051 GGCAAAACAA AGAATCTAAA GGAGAGACAG GGAAGAGAAG TAAGGAAGGT
2101 CATAGTTTGG AGATGGAGAA TGAGAATCTT GTAGAAAATG GTGCAGACTC
2151 CGATGAAGAT GATAACAGCT TCCTCAAACA ACAATCTCCA CAAGAACCCA
2201 AGTCTCTGAA TTGGTCGAGT TTTGTAGACA ACACCTTGGC TGAAGAATTC
2251 ACTACTCAGA ATCAGAAATC CCAGGATGTG GAACTCTGGG AGGGAGAAGT

```



```

2301 GGTCAAAGAG CTCTCTGTGG AAGAACAGAT AAAGAGAAAT CGGTATTATG
2351 ATGAGGATGA GGATGAAGAG TGACAAATTG CAATGATGCT GGGCCTTAAA
2401 TTCATGTTAG TGTTAGCGAG CCACTGCCCT TTGTCAAAAT GTGATGCACA
2451 TAAGCAGGTA TCCCAGCATG AAATGTAAAT TACTTGGAAG TAACCTTGGA
2501 AAAGAATTCC TTCTTAAAT CAAAAACAAA AAAAAAAAC AAAAAAACA
2551 CATTCTAAAT ACTAGAGATA ACTTTACTTA AATTCTTCAT TTTAGCAGTG
2601 ATGATATGCG TAAGTGCTGT AAGGCTTGA ACTGGGAAA TATTCCACCT
2651 GATAATAGCC CAGATTCTAC TGTATTCCCA AAAGGCAATA TTAAGGTAGA
2701 TAGATGATTA GTAGTATATT GTTACACACT ATTTTGAAT TAGAGAACAT
2751 ACAGAAGGAA TTTAGGGGCT TAAACATTAC GACTGAATGC ACTTTAGTAT
2801 AAAGGGCACA GTTTGTATAT TTTTAAATGA ATACCAATTT AATTTTTTAG
2851 TATTTACCTG TTAAGAGATT ATTTAGTCTT TAAATTTTT AGGTTAATTT
2901 TCTTGCTGTG ATATATATGA GGAATTTACT ACTTTATGTC CTGCTCTCTA
2951 AACTACATCC TGAACCTCGAC GTCCTGAGGT ATAATACAAC AGAGCACTTT
3001 TTGAGGCAAT TGA AAAACCA ACCTACACTC TTCGGTGCTT AGAGAGATCT
3051 GCTGTCTCCC AAATAAGCTT TTGTATCTGC CAGTGAATTT ACTGTACTCC
3101 AAATGATTGC TTTCTTTTCT GGTGATATCT GTGCTTCTCA TAATTACTGA
3151 AAGCTGCAAT ATTTTAGTAA TACCTTCGGG ATCACTGTCC CCCATCTTCC
3201 GTGTAGAGC AAAGTGAAGA GTTTAAAGGA GGAAGAAGAA AGAAGTGTCT
3251 TACACCACTT GAGCTCAGAC CTCTAAACCC TGTATTCCC TTATGATGTC
3301 CCGTTTTTGA GACACTAATT TTTAAATACT TACTAGCTCT GAAATATATT
3351 GATTTTTATC ACAGTATTCT CAGGGTGAAA TTAACCAAC TATAGGCCCT
3401 TTTCTGGGA TGATTTCTA GTCTTAAGGT TTGGGACAT TATAAAGTGG
3451 AGTACATTG TTGTACACAG TTGATATTCC AAATTGTATG GATGGGAGGG
3501 AGAGGTGCTT TAAGCTGTAG GCTTTTCTTT GTACTGCATT TATAGAGATT
3551 TAGCTTTAAT ATTTTTTAGA GATGTA AAC ATTCTGCTTT CTTAGTCTTA
3601 CCTAGTCTGA AACATTTT TCAATAAAG ATTTTAATTA AAATTTGAAA
3651 AAAAAA AAAA

```

BLAST Results

Entry HS512217 from database EMBL:
human STS SHGC-14654.
Length = 250
Minus Strand HSPs:
Score = 1202 (180.3 bits), Expect = 1.8e-46, P = 1.8e-46
Identities = 242/244 (99%)

Medline entries

95263566:
Three different rearrangements in a single intron truncate
sterol regulatory element binding protein-2 and produce
sterol-resistant phenotype in three cell lines. Role of introns
in protein evolution.

93258417:
Characterization of a pollen-specific cDNA from sunflower
encoding a zinc finger protein.

Peptide information for frame 1

ORF from 94 bp to 2370 bp; peptide length: 759
Category: similarity to known protein

```

1 MESSPFNRRQ WTSLSLRVTA KELSLVNKNK SSAIVEIFSK YQKAAEETNM
51 EKKRSNTENL SQHFRKGLT VLKKKWENPG LGAESHTDSL RNSSTEIRHR
101 ADHPPAEVTS HAASGAKADQ EEQIHPRSL RSPPEALVQG RYPHIKDGED
151 LKDHSTESKK MENCLGESRH EVEKSEISEN TDSAGKIEKY NVPLNRLKMM
201 FEKGEPTQTK ILRAQSRAS GRKISENSYS LDDLEIGPGQ LSSSTFDSEK
251 NESRRNLELP RLSETSIKDR MAKYQAAVSK QSSSTNYTNE LKASGGEIKI
301 HKMEQKENVP PGPEVCITHQ EGEKISANEN SLAVRSTPAE DDSRDSQVKS
351 EVQQPVHPKP LSPDSRASSL SESSPPKAMK KFQAPARETC VECQKTVYPM
401 ERLLANQVVF HISCFRCSYC NNKLSLGTYA SLHGRIYCKP HFNQLFKSKG
451 NYDEGFGHRP HKDLWASKNE NEEILERPAQ LANARETPHS PGVEDAPIAK
501 VGVLAASMEA KASSQKEKD KPAETKKLRI AWPPPTLGLS SGSALEEGIK
551 MSKPKWPPEP EISKPEVPED VDLDLKKLRR SSSLKERSRP FTVAASFQST
601 SVKSPKTVSP PIRKWSMSE QSEESVGGRV AERKQVENAK ASKKNGNVGK
651 TTWQNKESKG ETGKRSKEGH SLEMENENLV ENGADSDDED NSFLKQSQSQ
701 EPKSLNWSSF VDNTFAEEFT TQNKQSQDVE LWEGEVVKEL SVEEQIKRNR

```

751 YYDEDEDEE

BLASTP hits

Entry CG22818_1 from database TREMBL:
 "SREBP-2"; product: "mutant sterol regulatory element binding protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-2 (SREBP-2) mRNA, complete cds. Cricetulus griseus (Chinese hamster)
 Length = 839
 Score = 1502 (528.7 bits), Expect = 3.9e-154, P = 3.9e-154
 Identities = 290/380 (76%), Positives = 322/380 (84%)

Entry S28507 from database PIR:
 transcription factor SF3 - common sunflower
 Length = 219
 Score = 212 (74.6 bits), Expect = 6.3e-18, Sum P(2) = 6.3e-18
 Identities = 36/82 (43%), Positives = 55/82 (67%)

Entry NTLIMDOM_1 from database TREMBL:
 "SF3"; product: "LIM-domain SF3 protein"; N.tabacum mRNA for LIM-domain protein Nicotiana tabacum (common tobacco)
 Length = 189
 Score = 216 (76.0 bits), Expect = 1.0e-16, P = 1.0e-16
 Identities = 42/94 (44%), Positives = 57/94 (60%)

Alert BLASTP hits for DKFZphut1_18i19, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_18i19, frame 1

Report for DKFZphut1_18i19.1

[LENGTH] 759
 [MW] 85225.57
 [pI] 6.41
 [HOMOL] TREMBL:CG22818_1 gene: "SREBP-2"; product: "mutant sterol regulatory element binding protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-2 (SREBP-2) mRNA, complete cds. 1e-151
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR257w] 3e-05
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YGR162w TIF4631 - mRNA cap-binding protein] 1e-04
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YGR162w TIF4631 - mRNA cap-binding protein] 1e-04
 [BLOCKS] BL00478B
 [PIRKW] zinc finger 9e-16
 [PIRKW] DNA binding 9e-16
 [SUPFAM] LIM metal-binding repeat homology 9e-16
 [PROSITE] MYRISTYL 6
 [PROSITE] LIM_DOMAIN_1 1
 [PROSITE] AMIDATION 2
 [PROSITE] CAMP_PHOSPHO_SITE 4
 [PROSITE] CK2_PHOSPHO_SITE 28
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 15
 [PROSITE] ASN_GLYCOSYLATION 6
 [PFAM] LIM domain containing proteins
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 5.53 %

SEQ MESSPFNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRSNTENL
 SEG
 1ct1-
 SEQ SQHFRKGTTLTVLKKKWNPGGLGAESHTDSLRSSTEIRHRADHPPEVTSAAASGAKADQ
 SEG
 1ct1-
 SEQ EEQIHPRSLRSPPEALVQGRYPHIKDGEDLDKHSTESKKMENCLGESRHEVEKSEISEN
 SEG
 1ct1-
 SEQ TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ
 SEG

```

1ctl- .....
SEQ    LSSSTFDSEKNESRRNLELPRLSETSIKDRMAKYQAAVSKQSSSTNYTNELKASGGEIKI
SEG    .....
1ctl- .....

SEQ    HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKSEVQQPVHKKP
SEG    .....x
1ctl- .....

SEQ    LSPDSRASSLSESSPPKAMKKFQAPARETCVECQKTVYPMERLLANQQVFHISCFRCSYC
SEG    xxxxxxxxxxxxxxxx.....
1ctl- .....ETTTTETTTTCEEEETEEEEETTTTBTBT

SEQ    NNKLSLGTASYASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEEILERPAQ
SEG    .....
1ctl- TCBCBTBEEETETEEETTTTTTTTTTCTTTTCTTT.....

SEQ    LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQQEKEDKPAETKKLRIAWPPPELGS
SEG    .....
1ctl- .....

SEQ    SGSALEEGIKMSKPKWPPDEISKPEVPEDVDLDLKKLRSSSLKERSRPFTVAASFQST
SEG    .....xxxxxxxxxxxxxxxxxxxxx.....
1ctl- .....

SEQ    SVKSPKTVSPPIRKGWSMSEQSEESVGGRAERKQVENAKASKKNGNVGKTTWQNKESKG
SEG    .....
1ctl- .....

SEQ    ETGKRSKEGHSLEMENENLVENGADSDDDNSFLKQQSPQEPKSLNWSFVDNTFAEEFT
SEG    .....
1ctl- .....

SEQ    TQNQKSQDVELWEGEVVKELSVEEQIKRNRYYDEDEDEE
SEG    .....xxxxxxx
1ctl- .....

```

Prosite for DKFZphut1_18i19.1

PS00001	29->33	ASN_GLYCOSYLATION	PDOC00001
PS00001	59->63	ASN_GLYCOSYLATION	PDOC00001
PS00001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00001	286->290	ASN_GLYCOSYLATION	PDOC00001
PS00001	706->710	ASN_GLYCOSYLATION	PDOC00001
PS00004	52->56	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	65->69	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	222->226	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	579->583	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS00005	19->22	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	184->187	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	253->256	PKC_PHOSPHO_SITE	PDOC00005
PS00005	266->269	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	583->586	PKC_PHOSPHO_SITE	PDOC00005
PS00005	601->604	PKC_PHOSPHO_SITE	PDOC00005
PS00005	604->607	PKC_PHOSPHO_SITE	PDOC00005
PS00005	642->645	PKC_PHOSPHO_SITE	PDOC00005
PS00005	662->665	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	168->172	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	266->270	CK2_PHOSPHO_SITE	PDOC00006
PS00006	294->298	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	326->330	CK2_PHOSPHO_SITE	PDOC00006
PS00006	337->341	CK2_PHOSPHO_SITE	PDOC00006

PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00006	389->393	CK2_PHOSPHO_SITE	PDOC00006
PS00006	467->471	CK2_PHOSPHO_SITE	PDOC00006
PS00006	514->518	CK2_PHOSPHO_SITE	PDOC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS00006	658->662	CK2_PHOSPHO_SITE	PDOC00006
PS00006	686->690	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00006	709->713	CK2_PHOSPHO_SITE	PDOC00006
PS00006	714->718	CK2_PHOSPHO_SITE	PDOC00006
PS00006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS00007	223->230	TYR_PHOSPHO_SITE	PDOC00007
PS00007	222->230	TYR_PHOSPHO_SITE	PDOC00007
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	502->508	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	548->554	MYRISTYL	PDOC00008
PS00008	627->633	MYRISTYL	PDOC00008
PS00009	220->224	AMIDATION	PDOC00009
PS00009	662->666	AMIDATION	PDOC00009
PS00478	390->425	LIM_DOMAIN_1	PDOC00382

Pfam for DKFZphut1_18i19.1

HMM_NAME	LIM domain containing proteins		
HMM	*CagCNrpIyDREivMRAMNKvWHpECFrCcdCqqPLtegdeFYErDGRI		
	C C++++Y+ E++ A+ V+H++CFRC+ C+ L+ G+ + ++ GRI		
Query	390	CVECQKTVYPMERLL-ANQQVFHISCFRCSYCNKLSLGT-YASLHGRI	436
HMM	YCKhDYrrFg*		
	YCK+++ ++F+		
Query	437	YCKPHFNQLFK	447

DKFZphutel_18i4

group: uterus derived

DKFZphutel_18i4 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes.

weak similarity to C.elegans D2085.2

complete cDNA, complete cds, few EST hits

Sequenced by AGOWA

Locus: /map="7q31"

Insert length: 1568 bp

Poly A stretch at pos. 1551, polyadenylation signal at pos. 1523

```
1  GCCGAGCGGA GAGGGTAGAG ACGGGGTTC ACCGTGTTAG CCAAGATGGT
51 CTCGATCTCC TGACCTCGTG ATCCGCCCGC CTCGGCCTCC CAAAGTGGTG
101 GGATTACAGG CGTGAGCCAC TGCGCCCGGC CTGTTGTACA GTTATTAAAG
151 TTATCATTTA ACATGGAAGA AGATGAGTTC ATTGGAGAAA AAACATTCCA
201 ACGTTATTGT GCAGAATTCA TTAACATTC ACAACAGATA GGTGATAGTT
251 GGGAAATGGAG ACCATCAAAG GACTGTTCTG ATGGCTACAT GTGCAAAATA
301 CACTTTCAAA TTAAGAAATGG GTCTGTGATG TCACATCTAG GAGCATCTAC
351 CCATGGACAG ACATGTCTTC CCATGGAGGA GGCTTTCGAG CTACCCCTGG
401 ATGATTGTGA AGTGATTGAA ACTGCAGCAG CGTCCGAAGT GATTAAATAT
451 GAGTATCATG TCTTATATTC CTGTAGCTAC CAAGTGCCCTG TACTTTACTT
501 TAGGGCAAGC TTTTATAGATG GGAGACCTTT AACTCTGAAG GACATATGGG
551 AAGGAGTTCA TGAGTGCTAT AAGATGCGAC TGCTACAGGG ACCATGGGAC
601 ACTATTACGC AACAGGAACA TCCAATACTT GGGCAACCCT TTTTGTACT
651 TCATCCCTGC AAGACGAATG AATTCATGAC TCCTGTATTA AAGAATTCTC
701 AGAAAATCAA TAAGAAATGTC AACTATATCA CATCATGGCT GAGCATTGTA
751 GGGCCAGTTG TTGGGCTGAA TCTACCTCTG AGTTATGCCA AAGCAACGTC
801 TCAGGATGAA CGAAATGTCC CTTAACAAAG TTCTTCTATT GAGTTTAGGA
851 ATTGCGGCAC GAAGAATGCC AAGAGTTTAC CTGGCCAGCC CTGGCTTTAA
901 TAGGACTGAT ACCATGGAAT ATTCATCTC ACCAAGATGT GACATGGATT
951 ATTTTTCCTT TGGACACAAA TGTCTACAGC AACTGATGTT TGATAGGCTG
1001 AATGTTTAGA AGAAACACTT CAAAGGGATA CATCATGGCC AGGCATGGTG
1051 GCTCACACCT GTAATCCAAG CACTTTGGGA GGCCAAGGTG GGAGCATCAC
1101 TTGATCCTGG GAGTTCGAGA CCAGCCTGGG CAACATGGTG AAACCCTGTC
1151 GGTACAAAAA AATACAAAAA TTTGCCTGTT TATGGTGGTG TGTTCCTGTA
1201 GTCCAGCTC CCCAGGAGGC TGAGGTGGGA GGTGGCTTT AACCCAGGAG
1251 GCAGAGGTTG CAGTGAGCTG AGACTGTGCC ACTGCAGTCC AGCCTGGGTG
1301 ACAGAGCCAG ACCTGTCTC GGGAAAAAAA AAAAAAAA AAAGACACAT
1351 CACTATAAAT AGCAAAAAA CAAATCTAAC TTATTAATAC TAGGAATACC
1401 AACATTATTA GGGCACTTGC AGGTTATTCT TTTCTAGGCC AAGTACTTCA
1451 CTTCCATTG TCTGACATGG AGATTGAGGG AGAAATGTAT TTGTGTGTTT
1501 ATTTTAATGT AAGATATATA AAAATTAAAT TACTGGATT ACCTGTCCCT
1551 GAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 163 bp to 822 bp; peptide length: 220

Category: similarity to unknown protein

BLASTP hits

Alert BLASTP hits for DKFZphut1 18i4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphutet1 18i4, frame 1

Report for DKFZphutel 18i4.1

[BLOCKS]	BL00221E	
[PROSITE]	MYRISTYL	2
[PROSITE]	CK2_PHOSPHO_SITE	4
[PROSITE]	PKC_PHOSPHO_SITE	2
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha Beta	

```
SEQ      MEDEFIGEKTFQRYCAEFIKHSQQIGDSEWRPSKDCSDGYMCKIHQFIKNGSVMSHLG
PRD      cccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccc

SEQ      ASTHGQTCLPMEAEFELPLDDCEVIETAASAEVIKYEYHVLVYSCSYQVPVLVYFRASFLDG
PRD      cccccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccc

SEQ      RFLTLDKIWEGVHECYKMRLLQGPWDITITQEHPILGQFFVLVHPCKTNEFMTFVLKNSQ
PRD      cccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccc

SEQ      KINKNVNYITSWLSIVGPVVGGLNPLSYAKATSQDERNVCP
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphutel 18i4.1

PS00001	52->56	ASN_GLYCOSYLATION	PDOC00001
PS00005	124->127	PKC_PHOSPHO_SITE	PDOC00005
PS00005	179->182	PKC_PHOSPHO_SITE	PDOC00005
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00008	53->59	MYRISTYL	PDOC00008
PS00008	131->137	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1 18i4.1)

DKFZphut1_1811

group: nucleic acid management

DKFZphtes3_15j18 encodes a novel 184 amino acid protein with similarity to *S. cerevisiae* putative ribosomal protein YHR148w.

The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome subunit.

The new protein can find application in modulation of ribosome assembly, structure and function.

strong similarity to *S.cerevisiae* YHR148w

complete cDNA, complete cds, EST hits,
potential start at Bp 45 matches kozak consensus ANNAtgG
gene disruption of YHR148w is lethal!

Sequenced by AGOWA

Locus: unknown

Insert length: 1076 bp

Poly A stretch at pos. 1035, polyadenylation signal at pos. 1006

```
1 GCGCGCTCTC AGCTTCGGGT CCTGCGGCTG CGGCTGCCGC CATCATGGTG
51 CGGAAGCTTA AGTTCCACGA GCAGAAGCTG CTGAAGCAGG TGGACTTCCT
101 GAACTGGGAG GTCACCGACC ACAACCTGCA CGAGCTGCGC GTGCTGCGGC
151 GTTACCGGCT GCAGCGGCGG GAGGACTACA CGCGCTACAA CCAGCTGAGC
201 CGTGCCGTGC GTGAGCTGGC GCGGCGCCTG CGCGACCTGC CCGAACGCGA
251 CCAGTTCCGC GTGCGCGCTT CGGCCGCGCT GCTGGACAAG CTGTATGCTC
301 TCGGCTTGTT GCCCACGCGC GGTTCGCTGG AGCTCTGCGA CTTCTGTCAG
351 GCCTCGTCCT TCTGCCGCGG CCGCCTCCCC ACCGTGCTCC TCAAGCTGCG
401 CATGGCGCAG CACCTTCAGG CTGCCGTGGC CTTTGTGGAG CAAGGGCACC
451 TACGCGTGGG CCCTGACGTG GTTACCGACC CCGCCTTCCT TGTACGCGC
501 AGCATGGAGG ACTTTGTAC TGGGTGGAC TCGTCCAAGA TCAAGCGGCA
551 CGTGCTAGAG TACAATGAGG AGCGCATGA CTTCGATCTG GAAGCCTAGC
601 GGATCTCCCA CTTTGCATGG CTGTCTTTTA CAGATGGGAA AACTGAGGCC
651 TGATGCTGGA GATTCTATGA GGGTGCTCTC CTCAGGGGTA TCAGACGGTC
701 GTAGGTTCTT AAGAATTGA TTCATCAGTG GCAGGCCATG CATAGAGCCA
751 CGGGAGGTGC GTCCTTGTTT TCCAGGAAAT GTTCTTAGAA CTTGGACTAC
801 TGATTATTAA TTGACTGTGC CTTGGGAAAC AGTGGGAAGT AACTTGGTGC
851 AGCACTGGGG TATTGTTGGA CTGGTTCAAT TCGTTAACT CGAATTCTTG
901 CTCCTGGCCG TGGTTAAGCT GTGTACAGAT GATGGAGAGT TTGGCCTCAA
951 GTTTTATATA ACTGAGCGAG ACTAGTGTTC AGGATCTCCT CCCTTGTTTA
1001 AATGTCAATA AATGCCCAA CTGCTTTGTA AGCTCAAAA AAAAAAAAAA
1051 AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 45 bp to 596 bp; peptide length: 184

Category: strong similarity to known protein

```
1 MVRKLKFEHQ KLLKQVDFLN WEVTDHNLHE LRLRRYRLQ RREDYTRYNQ
51 LSRVRELAR LRLDLPERDQ FRVRASAALL DKLYALGLVP TRGSLELCDF
101 VTASSFCRRR LPTVLLKLRM AQHLQAAVAF VEQGHVVRVGP DVVTDPAFLV
151 TRSMEDEFVT VDSSKIKRHV LEYNEERDDF DLEA
```

BLASTP hits

Pedant information for DKFZphutel 1811, frame 3

457


```
          ++++++ +          +++++W++ S+          ++R+ + Y+ +
Query      147 AFLVTRS---M-----EDFVTWVDSSK-----IKRHVLEYNEERD . 178
HMM          rIIEReWiplkINELLVVEY*
          +++ +
Query      179 DFDLE----- 183
```

DKFZphutel_19f19

group: transmembrane protein

DKFZphutel_19f19 encodes a novel 204 amino acid protein with similarity to murine p24 protein.

Murine p24 is expressed only in brain where it is localized exclusively in neurons. It seems to be a neuron-specific membrane protein localised in intracellular organelles of highly differentiated neural cells and may play a role in the neural organelle transport system. As p24, the novel protein contains 2 transmembrane regions, but it contains not the sequence homologous to the microtubule-binding domain of microtubule-associated proteins present in p24.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to mouse P24 protein ;
membrane regions: 2
Summary DKFZphutel_19f19 encodes a novel 204 amino acid protein, with
similarity to mouse P24 protein.

similarity to mouse P24 protein

complete cDNA, complete cds, EST hits,
2 TM-domains

Sequenced by AGOWA

Locus: /map=14.8 cR from top of Chr20 linkage group

Insert length: 2042 bp

Poly A stretch at pos. 1958, polyadenylation signal at pos. 1940

```
1 GCAGGCAGAG AGATGAGGAA ACTGAGACCC AGAAAGGTGG AAGCACTTGT
51 CTAAGGTCAC GCCTCCAGGA AGCAGTGTGT CCACGACTCC AGTCCAAGTG
101 GTCAGGCTCC AGAGCCCACA GTCCCAGGGG TCCATGATGC CGAGCTGCAA
151 TCGTTCCTGC AGCTGCAGCC GCGGCCCCAG CGTGGAGGAT GGCAAGTGGT
201 ATGGGGTCCG CTCCTACCTG CACCTCTTCT ATGAGGACTG TGCAGGCACT
251 GCTCTCAGCG ACGACCCTGA GGGACCTCCG GTCCTGTGCC CCCGCCGGCC
301 CTGGCCCTCA CTGTGTTGGA AGATCAGCCT GTCCTCGGGG ACCCTGCTTC
351 TGCTGCTGGG TGTGGCGGCT CTGACCACTG GCTATGCAGT GCCCCCCAAG
401 CTGGAGGGCA TCGGTGAGGG TGAGTTCCTG GTGTTGGATC AGCGGGCAGC
451 CGACTACAAC CAGGCCCTGG GCACCTGTCT CCTGGCAGGC ACAGCGCTCT
501 GTGTGGCAGC TGGAGTTCTG CTCGCCATCT GCCTCTTCTG GGCCATGATA
551 GGCTGGCTGA GCCAGGACAC CAAGGCAGAG CCCTTGGACC CCGAAGCCGA
601 CAGCCACGTG GAGGTCTTCG GGGATGAGCC AGAGCAGCAG TTGTCAACCA
651 TTTTCCGCAA TGCCAGTGGC CAGTCATGGT TCTCGCCACC CGCCAGCCCC
701 TTTGGGGAAT CTTCTGTGCA GACTATCCAG CCCAAGAGGG ACTCCTGAGC
751 TGCCCAACAT GCCTAAGATG TGGGTCTTGG ATCCTTCCCC CTTCTCACCA
801 TAACCCCTCT TCAGTGTTCG CCCAACTTCT CCCTTTAGAG CCCAACTCCA
851 GGTCAAAATCT GGAGCTCAAA TCCCAGTGCT CCCTCCCCAG GAGTGGGGCC
901 CCAACTCTTC CAAGATACCA GCATTCTCTA AGTCCTCCCA AAACCTCCTA
951 CCCACACCCT CTTCCCAAGG CCCTCAGGGG CAGAAAACAT CTCCTTCAAC
1001 CCGTCCCCAC TCCTTCCTCT GCATGACCTT GGGCAAACCC TTGCCCTTTC
1051 AAGCCATCAG CTCCTGCCTC TCTGCCATGA GGGCTTTGGA TCAGATTCTT
1101 CTTCTCGCCA GGATGAGGAC ACGCACTGCC CTCCATAGAC ACAGATGAAG
1151 GGGTGGGGGT CATTGAGCTC GAATGGGTCC CAGATGCTCA CTTGGCCTTT
1201 CCCTGCAGGA TGAGTGAAGA CGTTTGCCTC TCACAGTGTG TCTTCTACCT
1251 GCATTTTGGC ATCAGAGCCC CCCAGCCCAC CCACCACAGG CAATTACTAG
1301 CCCTAGTTGA TAGGTGAGGT GGGTGAAGAA GGCTGGAGGT GACATGTCCG
1351 AGGTACACCA ACAGAGCAGC ATGCAGGAAC TAGAAACACA TCTTCAGCCT
1401 CCTCCTGGGC CAGCTCTTGT GCTACAGGTG GGGCGGAGCC AGCCCCCTAC
1451 CTTCTGTGTT CCCTGAGGGT CCTCAGGGTG GAGGACAGGT TTGGCCGAGA
1501 AAGACTAGCC AGAGGCCTGA TGGTCCAGG TGGCTCTGGA TATACTTTGG
1551 ATATGGATTT AAATGGTCTC TAAGAGCCGG GGGTAGGGGG CAGGAAAAAT
1601 GGGTTGTCTT TGCCCTCAA AGTCCACCTA CCTAGAAACC AAGCCACCGG
1651 TCTTGGCCGT GACCCTGATA ATAAATGGGC TCTCTCAGAG GCGCCAGCCC
1701 CTCCCTCCCC AGCCGGAGGC GTCATCTCTC TTCTGTACCA CTAGAGGGAG
1751 CTCTGATGCA GCTGGAGAGC AGCGCTCAAG GCTCTCGCCC CTCCCTCCCC
1801 TAACCCCTTAC CTTCAGTCTC CACCAGCCTG AAGGGCCTCC TAGGGGATCC
1851 TCAGCGGGCC CCCACCAGGG CACACCCTAC TGTCCTTTGT CCTCAGCCCC
1901 CCTCCTCATC CTGCACCCCT TCCATCCCAC CTTCCCTTTC AATAAACAGC
1951 TGGGATGGAA AAAAAAAAAA AGAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2001 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

Entry HS417348 from database EMBL:
 human STS WI-14697.
 Length = 290
 Minus Strand HSPs:
 Score = 1254 (188.2 bits), Expect = 3.0e-50, P = 3.0e-50
 Identities = 262/273 (95%)

Medline entries

97334404:
 A newly identified membrane protein localized exclusively in
 intracellular organelles of neurons.

Peptide information for frame 2

ORF from 134 bp to 745 bp; peptide length: 204
 Category: similarity to known protein

1 MPPSCNRSCS CSRGPSVEDG KWYGVRSYLH LFYEDCAGTA LSDDPEGPPV
 51 LCPRRWPSSL CWKISLSSGT LLLLLGVAAL TTGYAVPPKL EGIGEGEFLV
 101 LDQRAADYNQ ALGTCRLAGT ALCVAAGVLL AICLFWAMIG WLSQDTKAEP
 151 LDPEADSHVE VFGDEPEQQL SPIFRNASGQ SWFSPPASPF QSSSVQTIQP
 201 KRDS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19f19, frame 2

TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein,
 complete cds., N = 1, Score = 295, P = 3.8e-26

>TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein,
 complete cds.
 Length = 196

HSPs:

Score = 295 (44.3 bits), Expect = 3.8e-26, P = 3.8e-26
 Identities = 58/139 (41%), Positives = 81/139 (58%)

Query: 2 MPSCNRSCSCSRGPSVEDGKW---YGVRSYLHLFYEDCAGTALSDDPEGPPVLCPRRWP 58
 M SC+ +C R + +G + YGVRSYLH FYEDC + + + P R W
 Sbjct: 1 MTSCSNTCGSRRAQADTEGGYQQRVGVRSYLHQFYEDCTASIWEYEDDFQIQRSPNR-WS 59

Query: 59 SLWCISLSSGTL LLLLLGVAALTTGYAVPPKLEGIGEGEFLVLDQRAADYNQALGTCRLA 118
 S+ WK+ L SGT+ ++LG+ L G+ VPPK+E GE +F+V+D A YN AL TC+LA
 Sbjct: 60 SVFWKVLISGTVFVILGLTVLAVGFLVPPKIEAFGEADFMVVDTHAVKYNALDTCCLA 119

Query: 119 GTALCVAAGVLLAICLFWAM 138
 G L G +A CL ++
 Sbjct: 120 GAVLFCIGTSMAGCLLSV 139

Pedant information for DKFZphut1_19f19, frame 2

Report for DKFZphut1_19f19.2

[LENGTH] 204
 [MW] 21983.07
 [pI] 4.69
 [HOMOL] TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein, complete
 cds. 7e-19
 [PROSITE] MYRISTYL 4

```

[PROSITE]    CAMP_PHOSPHO_SITE      1
[PROSITE]    CK2_PHOSPHO_SITE       3
[PROSITE]    PKC_PHOSPHO_SITE       1
[PROSITE]    ASN_GLYCOSYLATION      2
[KW]         TRANSMEMBRANE 2
[KW]         LOW_COMPLEXITY 10.29 %

```

```

SEQ  MMPSCNRSCSCSRGPSVEDGKWYGVRSYLHLFYEDCAGTALSDDPEGPPVLCPRRPWPSL
SEG  .....
PRD  cccccccccccccccccccccceehhhhhccccccccccccccccccccccccccccce
MEM  .....MM

```

```

SEQ  CWKISLSSGTLTLLLLGVAALTGTGYAVPPKLEGIGEGEFLVLDQRAADYNQALGTCRLAGT
SEG  ....xxxxxxxxxxxxxxxxxxxxxx
PRD  eeeeeccccceccccceccccccccccccccccceccccccccchhhhhhhchh
MEM  MMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMM

```

```

SEQ  ALCVAAGVLLAICLFWAMIGWLSQDTKAEPLDPEADSHVEVFGDEPEQQLSPIFRNASGQ
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhcccccccccccccecccccccccccccccccccc
MEM  MMMMMMMMMMMMMMMMMMMMMMMMM.....

```

```

SEQ  SWFSPPASPFQSSVQTIQPKRDS
SEG  .....
PRD  ccccccccccccccecccccccc
MEM  .....

```

Prosites for DKFZphut1_19f19.2

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	176->180	ASN_GLYCOSYLATION	PDOC00001
PS00004	201->205	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	114->117	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	157->161	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	119->125	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_19f19.2)

DKFZphute1_19g19

group: uterus derived

DKFZphute1_19g19 encodes a novel 400 amino acid protein, with strong but partial similarity to a bovine elastin-related protein expressed in fetal calf ligamentum nuchae.

The novel protein contains 2 RGD cell attachment sites.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to bovine elastin fragment

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: map=54.9 cR from top of Chr3 linkage group

Insert length: 3244 bp

Poly A stretch at pos. 3227, polyadenylation signal at pos. 3216

```
1  GTAACGTCAG TAAGTCCCGC TTGGCCCTGG AGTCCACGCG GATTTTCGAA
51  GCTGGGGGCTG GCAAGAGGCC GCTGGACACC ACGCTCCAGT CGTCAGCCCA
101 CTTCTTAGCT GAACAGCGCG AGGCGGCGCG AGCGAGCCGG GTCCCAACCAT
151 GGCCGCGAAT TATTCCAGTA CCAGTACCCG GAGAGAACAT GTCAAAGTTA
201 AAACCAAGCTC CCAGCCAGGC TTCTTGAAC GGCTGAGCGA GACCTCGGGT
251 GGGATGTTTG TGGGGCTCAT GGCCTTCTCT CTCTCTCTCT ACCTAATTTT
301 CACCAATGAG GGCCGCGCAT TGAAGACGGC AACCTCATTG GCTGAGGGGC
351 TCTCGCTTGG GGTGTCTCCT GACAGCATCC ACAGTGTGGC TCCGGAGAAT
401 GAAGGAAGGC TGGTGCACAT CATTGGCGCC TTACGGACAT CCAAGCTTTT
451 GTCTGATCCA AACTATGGGG TCCATCTTCC GGCTGTGAAA CTGCGGAGGC
501 ACGTGGAGAT GTACCAATGG GTAGAAACTG AGGAGTCCAG GGAGTACACC
551 GAGGATGGGC AGGTGAAGAA GGAGACGAGG TATTCCTACA AACTGAATG
601 GAGGTCAGAA ATCATCAACA GCAAAACTT CGACCGAGAG ATTGGCCACA
651 ATAACCCAG TGCCATGGCA GTGGAGTCAT TCACGGCAAC AGCCCCCTTT
701 GTCCAAATTG GCAGGTTTTT CCTCTCGTCA GGCTCATCG ACAAGTCGA
751 CAACCTCAAG TCCCTGAGCC TATCCAAAGT GGAGGACCCT CATGTGGACA
801 TCATTGCGCG TGGAGACTTT TTCTACCACA GCGAAAATCC CAAGTATCCA
851 GAGGTGGGAG ACTTGCCTGT CTCCTTTTCC TATGCTGGAC TGAGCGGCGA
901 TGACCTTGAC CTGGGCCAG CTCACGTGGT CACTGTGATT CCGCGGCAGC
951 GGGGTGACCA GCTAGTCCCA TTCTCCACCA AGTCTGGGGA TACCTTACTG
1001 CTCCTGCACC ACGGGGACTT CTCAGCAGAG GAGGTGTTTC ATAGAGAACT
1051 AAGGAGCAAC TCCATGAAGA CCTGGGCGCT GCGGGCAGCT GGCTGGATGG
1101 CCATGTTTAT GGGCCTCAAC CTTATGACAC GGATCCTCTA CACCTTGGTG
1151 GACTGTGTTT CTGTTTTCCG AGACCTGGTC AACATTGGCC TGAAGCCTT
1201 TGCTTCTGT GTGGCCACCT CGCTGACCTT GCTGACCGTG GCGGCTGGCT
1251 GGTCTCTCTA CCGACCCCTG TGGGCCCTCC TCATTGCCGG CCTGGCCCTT
1301 GTGCCATACC TTGTTGCTCG GACACGGGTG CCAGCCAAAA AGTTGGAGTG
1351 AAAAGACCCCT GGCACCCGCG CGACACCTGC GTGAGCCCTA GGATCCAGGT
1401 CCTCTCTCAC CTCTGACCCA GCTCCATGCC AGAGCAGGAG CCGCGGTCAA
1451 TTTTGGACTC TGCACCCCTT CTCTCTTCA GGGGCCAGAC TTGGCAGCAT
1501 GTGACACAGG TTGGTGTCTA CCAGCTCATG TCTTCCCCAC ATCTCTCTT
1551 GCCAGTAAGC AGCTTTGGTG GGCAGCAGCA GCCATGAATG GCAAGCTGAC
1601 AGCTTCTCCT GCTGTTTCCT TCCTCTCTTG GACTGAGTGG GTACGGCCAG
1651 CCACTCAGCC CATTGGCAGC TGACAACGCA GACACGCTCT ACGGAGGCCT
1701 GCTGATAAAG GGCTCAGCCT TGCCGTGTGC TGCTTCTCAT CACTGCACAC
1751 AAGTGCCATG CTTTGCCACC ACCACCAAGC ACATCTGTGA TCCTGAAGGG
1801 CGGCCGTTAG TCATTACTGC TGAGTCTCTG GTCAACGACA GACACACTGG
1851 GCATGGACCC CTCAAAAGCAG GCACACCCAA AACACAAGTC TGTGGCTAGA
1901 ACCTGATGTG GTGTTTAAAA GAGAAGAAAC ACTGAAGATG TCCTGAGGAG
1951 AAAAGCTGGA CATATACTGG GCTTCAACT TATCTTATGG CTTGGCAGAA
2001 TCTTTGTAGT GTGTGGGATC TCTGAAGGCC CTATTTAAGT TTTTCTCTGT
2051 TACTTTGCTG CTTTCATGTG ACTTCTCTAC CCCAAGAGGA AGTTTTCTGA
2101 AATAAGATTT AAAAAACAAA CAAAAAAAC ACTTAATATT TCAGACTGTT
2151 ACAGGAAACA CCCTTTAGTC TGTCAAGTTA ATTCAAGACA CTGAAAGGTG
2201 TTAATTTGGG GTATGTGGTT TGATTGATAA AAAGTTACCT CTCAGTATTT
2251 TGTGTCACTG AGAAGCTTTA CAATGGATGC TTTTGAACA AGTATCAGCA
2301 AAAGGATTTG TTTTCACTCT GGGAGGAGAG GGTGGAGAAA GCACTTGCTT
2351 TCATCTCTCT GCATCGGAAA CTCCCTATG CACTTGAAGA TGGTTTAAAA
2401 GATTAAAGAA ACGATTAAGA GAAAAGGTTG GAAGCTTTAT ACTAAATGGG
2451 CTCCTTCATG GTGACGCCCC GTCAACCACA ATCAAGAAGT GAGGCCTGAG
2501 GCTGGTGTGA CAATGCCACG GCCTGCCTGG CTGCTTTCAC CTGGGAGTGC
2551 TTTTCGATGT GGCACCTGGG CTTCTAGGG CTGCTTCTGA GTGGTCTTTT
2601 CACGTGTTGT GTCATAGCT TTAGTCTTCC TAAATAAGAT CCACCCACAC
```

```

2651 CTAAGTCACA GAATTCTTAA GTTCCCAAC TACTCTCACA CCCTTTTAAA
2701 GATAAAGTAT GTTGTAAACCA GGATGTCTTA AATGATTCTT TGTGTACCTT
2751 TTCTGTCATA TTCAGAAACC GTTTGTGCC TGCTGGGAGT AATTCCTTTA
2801 GCAATTAAGT ATTTGGTAGC TGAATAAGGG GTCAGAACTT CTGAAACCAG
2851 AGATCTGTAA TCATCTCTAT TGGCCTGGGG TGCCTGTGCT ATAAATGAGT
2901 TTCTTCACAT GAAAAACACA GCCAGCCCAA GATGACTTAT CTGGGTTTAG
2951 GATTCAATAG TATTCATAA CTGCTTATTA CATGAGCAAT TTCATCAAAT
3001 CTCCAAACTC TTAAAGGATG CTTTCGGAAA ACACGCTGTA TACCTAGATG
3051 ATGACTAAAT GCAAAATCCT TGGGCTTTGG TTTTTTCTA GTAAGGATT
3101 TAAATAACTG CCGACTTCAA AAGTGTCTT AAAACGAAAG ATAATGTTAA
3151 GAAAAATTG AAAGCTTTGG AAAACCAAAT TTGTAATATC ATTGTATTTT
3201 TTATTTAAAG TTTTGAATA AATTCTTAAA AAAAAAAAAA AAAA

```

BLAST Results

```

Entry HS545355 from database EMBL:
human STS WI-14815.
Length = 436
Minus Strand HSPs:
Score = 2040 (306.1 bits), Expect = 6.2e-86, P = 6.2e-86
Identities = 420/426 (98%)

```

```

Entry HS932147 from database EMBL:
human STS WI-8531.
Length = 341
Minus Strand HSPs:
Score = 1705 (255.8 bits), Expect = 4.7e-70, P = 4.7e-70
Identities = 341/341 (100%)

```

Medline entries

```

86051793:
Bovine elastin cDNA clones: evidence for the occurrence of a
new elastin-related protein in fetal calf ligamentum nuchae.

```

Peptide information for frame 2

```

ORF from 149 bp to 1348 bp; peptide length: 400
Category: similarity to known protein

```

```

1 MAANYSSSTST RREHVKKVTS SQPGFLERLS ETSGGMFVGL MAFLLSFYLI
51 FTNEGRALKT ATSLAEGLSL VVSPDSIHSV APENEGRLVH IIGALRTSKL
101 LSDPNYGVHL PAVKLRRHVE MYQWVETES REYTEDGQVK KETRYSYNT
151 WRSEIINSKN FDREIGHNNP SAMAVESFTA TAPFVQIGRF FLSSGLIDKV
201 DNFKSLSLSK LEDPHVDIIR RGDFFYHSEN PKYPEVGDLR VSFSYAGLSG
251 DDPDLGPAHV VTVIARQRGD QLVPFSTKSG DTLILLHHGD FSAEEVFHRE
301 LRSNSMKTWG LRAAGWMAMF MGLNLMTRIL YTLVDWFPVF RDLVNIGLKA
351 FAFCVATSLT LLTVAAGWLF YRPLWALLIA GLALVPILVA RTRVPAKKLE

```

BLASTP hits

```

Entry I45887 from database PIR:
elastin - bovine (fragment)
Length = 40
Score = 131 (46.1 bits), Expect = 4.9e-08, P = 4.9e-08
Identities = 31/41 (75%), Positives = 34/41 (82%)

```

Alert BLASTP hits for DKFZphut1_19g19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut1_19g19, frame 2

Report for DKFZphut1_19g19.2

[LENGTH] 400

```

[MW]          44831.53
[pI]          7.23
[HOMOL]       PIR:I45887 elastin - bovine (fragment) 1e-06
[PROSITE]     RGD      2
[PROSITE]     MYRISTYL  3
[PROSITE]     CAMP_PHOSPHO_SITE  1
[PROSITE]     CK2_PHOSPHO_SITE   6
[PROSITE]     TYR_PHOSPHO_SITE   2
[PROSITE]     PKC_PHOSPHO_SITE   5
[PROSITE]     ASN_GLYCOSYLATION  1
[KW]          TRANSMEMBRANE 4

```

```

SEQ  MAANYSSSTSTRREHVKVKTSSQPGFLERLSETSGGMFVGLMAFLLSFYLIFTNEGRALKT
PRD  ccccecccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

```

```

SEQ  ATSLAEGLSLVSPDSIHSVAPENEGRLVHIIGALRTSKLLSDPNYGVHLPVAVKLRRHVE
PRD  hhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

```

```

SEQ  MYQWVETESREYTEDGQVKKETRYSYNTEWRSEIINSKNFDREIGHNPNPSAMAVESFTA
PRD  hheehhhhhhecccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....M

```

```

SEQ  TAPFVQIGRFFLSSGLIDKVDNFKSLSLSKLEDPHVDIIRRGDFYHSENPKYPEVGDLR
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

```

```

SEQ  VSFYSAGLSGDDPDLGPAHVVTVIARQRGDQLVPFSTKSGDTLLLLHHGDFSAAEVFHRE
PRD  ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

```

```

SEQ  LRSNSMKTWGLRAAGWMAMFMGLNLMTRILYTLVDWFPVFRDLVNIGLKAFACVATSLT
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

```

```

SEQ  LLTVAAGWLFYRPLWALLIAGLALVPILVARTRVPAKKE
PRD  hhhhccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

```

Prosites for DKFZphut1_19g19.2

PS00001	4->8	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	276->279	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	332->336	CK2_PHOSPHO_SITE	PDOC00006
PS00007	220->227	TYR_PHOSPHO_SITE	PDOC00007
PS00007	99->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00016	221->224	RGD	PDOC00016
PS00016	268->271	RGD	PDOC00016

(No Pfam data available for DKFZphut1_19g19.2)

DKFZphut1_19g22

group: cell structure and motility

DKFZphut1_19g22 encodes a novel 390 amino acid protein with very strong similarity to tuftelin/enamelin.

Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix.

The new protein can find application in modulation of tissue-calcification, especially the uterus.

complete cDNA, complete cds start at Bp 51, EST hits in 3' UTR,
human homolog of mouse tuftelin
tuftelin is described as a matrix protein of teeth but it seems also
to be present in the uterus matrix

Sequenced by AGOWA

Locus: unknown

Insert length: 3110 bp

Poly A stretch at pos. 3093, polyadenylation signal at pos. 3071

```

1 GCAGACACGCG GGGTGGACAA GTGGCGTGTG TGCTGCGACC CCGAGGGAAG
51 ATGAACGGGA CGCGGAACGT GTGTACCTGT GTGGACGTGC ACCCAGAGGA
101 CCAGGCGGCG GGCAGCGTGG ACATTCTCAG GCTGACTCTC CAGGGTGAAC
151 TGACAGGAGA TGAACCTGAA CACATAGCCC AGAAGGCGGG CAGGAAGACC
201 TATGCCATGG TGTCCAGCCA CTCAGCTGGT CATTCTCTGG CTTCAGAACT
251 GGTGGAGTCC CATGATGGAC ATGAGGAGAT CATTAAAGTG TACTTGAAGG
301 GGAGGTCTGG AGACAAGATG ATTCACGAGA AGAATATTAA CCAGCTGAAG
351 AGTGAAGTCC AGTACATCCA GGAGGCCAGG AACTGCCTAC AGAAGCTCCG
401 GGAGGATATA AGTAGCAAGC TTGACAGGAA CCTAGGAGAT TCTCTCCATC
451 GACAGGAGAT ACAGGTGGTG CTAGAAAAGC CAAATGGCTT TAGTCAGAGT
501 CCCACAGCCC TGTACAGCAG CCCACCTGAG GTGGACACCT GTATAAATGA
551 GGATGTTGAG AGCTTGAGGA AGACGGTGCA GGACTTGCTG GCCAAGCTTC
601 AGGAGGCCAA GCGGCAACAC CAGTCAGACT GTGTGGCTTT TGAGGTCACA
651 CTCAGCCGGT ACCAGAGGGA AGCAGAACAA AGTAATGTGG CCCTTCAGAG
701 AGAGGAGGAC AGAGTGGAGC AGAAAGAGGC AGAAGTCGGA GAGCTGCAGA
751 GCGCTTGCT AGGGATGGAG ACGGAGCATC AGGCCTTACT GCGGAAAGTG
801 AGGGAAGGGG AGGTGGCCCT AGAGGAACCT CGGAGCAACA ATGCTGACTG
851 CCAAGCAGAA CGAGAAAAGG CTGCTACCCT GGAAAAGGAA GTGGCCGGGT
901 TCGGGAGGAA GATCCACCAC TTGGATGACA TGCTCAAGAG CCAGCAGCGG
951 AAAGTCCGGC AAATGATAGA GCAGCTCCAG AATTCAAAG CTGTGATCCA
1001 GTCAAAGGAC GCCACCATCC AGGAGCTCAA GGAGAAAATC GCCTATCTGG
1051 AGGCAGAGAA TTTAGAGATG CATGACCGGA TGAACACCT GATAGAAAAA
1101 CAAATCAGTC ATGGCAACTT CAGCACCCAG GCCCGGGCCA AGACAGAGAA
1151 CCCGGGCACT ATTAGGATAT CCAAGCCGCC TAGCCCCAAG CCCATGCCCTG
1201 TCATCCGAGT GGTGGAAACC TGAGCTGCCT GGAGATGGTT GCTGCCATTG
1251 CTGCTGCCTC TGCTCGGAG AAGCCCACTG CCCCTGTGTT CTGTTAACAC
1301 TGCTTTGAC TTCCTGACTG TCCCTGGCT GCACCCAGGA CTTCGGGCTC
1351 CTGTGTCTCA CCATTCCCAA GCCCCTGGCC ACTCTAAGCT GGGCAGACGG
1401 AGCAGGAGCA CCTATTCAAG GCACTGCAGC CCTTTGGAAG ACATTGTCTT
1451 GCAAGCAGGA GCCAGGGCAA TATCTATATT CCTACAGTGA CTATTTTCT
1501 CTGTAGAGAG CCTCCCTTCT GTTGTAGACT GGACTCTGGC TCGGCCATAA
1551 GCCAGGCCCT CATCAGATTG GGAGAGGTGA CAAGATTGTC CTCAGCCCTA
1601 AAAGCTGGAG ACACAGATGT CCAGAGTGAT TGGAGAAATG CCTGGGGGAA
1651 TGAAGTTCCT TCCACAAACA CAGCTCAGTT CTTAGCAACA AACTGTTTGT
1701 TTTTCTACTT GCTCCATCTG CAGCCTACGC TGCCCTGGCC TCCTGCAGAC
1751 AGATAGTGGG GTTACCTGGC AAGGCCCTGT GAGAGCCAGT GAACCTAAGC
1801 TTTGACTGGG TGGCCTTGTC TTTCTGGGGA GGAGGGAATG TACATTCAAG
1851 GAGTAGCCTT TTGCGGAAAA ATTCTCTAGG GCTACAGACA GTCATGTGTG
1901 ACTTCTCTCT GCTGTGAAAA CTCCAGAGT CTCTTTAGGG ATTTTCCCTA
1951 AGGTGTACCA CCAGGCACAC CTCAGTCTTC TTGACCCAGA GCCTGAAAAC
2001 TGTTTCACT GGGTTCACAC AGTCCCAGCA AAATCCTCTT TGTATTTATT
2051 TTGCTAAGTT ATTGGTGGTT TTGCTTACAT CTCATGATTG ATATAATACC
2101 AAAGTTCTAT AGCCTTCTCT TGCAGTATTT GGATTTGCTT GAAACCGGGA
2151 AAAGTGTTC CATTAGGCTT GTTAATGTCA GAGTGACACT ATTATGAATC
2201 TTTCTCTCCC TTTCTCTGCG CTGTTTCTTC TCTCTTCTC CTTCAAACTT
2251 GCTCTGCAGC TAAGGAAGGT GAGTCTACTT TCCCTGAGGC TTTGGGGTCA
2301 GAGTATATGT TGTTTGAGA AAGAGGGCAA TCAGGACTCT TCTGGGACCC
2351 AGATGAGTTC TTCACTAGCC CTCTGAACC CCTTGCTCCA TAATTGGTCT
2401 TTTATCTTGG CTCTGAATGA CCCTGCAGGT CATCATGGTT TTCTTTTTTT
2451 ATTGTTTTTT TTTTTTCTG AGACAGAGTC TCACTCTGTC ACCCAGGCTG
2501 GAGTGCAGTG GCGCATCTC AGCTCACTGC AACCTCTGCC TCCCGGATTT
2551 AAGCGATTCT TCTGCCTCAG CCTCCCGAGT AGCTGGGACT ACAGGTGTGC

```



```

2601 CACCACGCCCT GGCTGATTTT TGTATTTTTA GTAGAGATGG GGTTCACCA
2651 TACTGGCTAG GCTGGTCTCG AATTCCTGAC CTCAGGTGAT CCACCCACCT
2701 CGGCTTCCCA AAGTGCTAGG ATTATAGGCT TGAGCTACTG TGCCCCGGCCC
2751 ATGGTGTTTT TCTTTAGGGC TCTTCCTACA GCCTTGAGAA GTAGATAGGC
2801 ATCAGAGTAT GGTACTATAG GAATCAGAAA AATTCAAAAC AAATGTGGAT
2851 TAAGTGTTTA GGCTCTATGT GGCTCACGCA GCCAGAATCC TTAAGTCTGT
2901 GTGTTTCTGT GTCTCAAGAC TGGGCTCACA TTCTGGCTTT GTCCATAACA
2951 ATGCTCTGGG ATTTCAAGGA GTTCCCTCAT TTGTAAAATG AGGGGGTCAG
3001 AGCAGGTGAT ATCCATGTTT CTTCCTTTTC TGATATTGTT GTCTGTGGCA
3051 TATTCTTTGT ATGGCGAATT TAATAAATTA TATTAATGTG TCTAAAAAAA
3101 AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

```

98200312:
Tuftelin--aspects of protein and gene structure

97228909:
Timing of the expression of enamel gene products during mouse tooth
development.

91340750:
Sequencing of bovine enamelin ("tuftelin") a novel acidic enamel
protein.

```

Peptide information for frame 3

ORF from 51 bp to 1220 bp; peptide length: 390
 Category: strong similarity to known protein

```

1 MNGTRNWCTL VDVHPEDQAA GSVDIRLRLTL QGELTGDELE HIAQKAGRKT
51 YAMVSSHSAG HSLASELVES HDGHEEIIKV YLKGRSGDKM IHEKNINQLK
101 SEVQYIQEAR NCLQKLREDI SSKLDRNLGD SLHRQEIQVV LEKPNQFSQS
151 PTALYSSPPE VDTICINEDVE SLRKTVDQDL AKLQEAQRQH QSDCVAFEV
201 LSRVQREAEQ SNVALQREED RVEQKEAEVG ELQRRLLGME TEHQALLAKV
251 REGEVALEEL RSNNADCQAE REKAATLEKE VAGLREKIH LDDMLKSQQR
301 KVRQMIEQLQ NSKAVIQSKD ATIQELKEKI AYLEAENLEM HDRMEHLIEK
351 QISHGNFSTQ ARAKTENPGS IRISKPPSPK PMPVIRVVET

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19g22, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphut1_19g22, frame 3

Report for DKFZphut1_19g22.3

```

[LENGTH]      390
[MW]           44264.09
[pI]           5.68
[HOMOL]        TREMBL:AF047704_1 product: "tuftelin"; Mus musculus tuftelin mRNA, complete
cds. 0.0
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w]
2e-11
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 2e-11
[FUNCAT]       1 genome replication, transcription, recombination and repair [M.
jannaschii, MJ1643] 7e-11
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 1e-08
[FUNCAT]       03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 6e-08
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YGL086w] 6e-08
[FUNCAT]       03.13 meiosis [S. cerevisiae, YNL250w] 7e-08

```

[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 7e-08
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
 repair) [S. cerevisiae, YKR095w] 1e-07
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR285w] 2e-07
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-07
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 1e-05
 [FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YNL243w]
 1e-04
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 -
 myosin-1 isoform] 4e-04
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-04
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 4e-04
 [FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YMR294w] 7e-04
 [EC] 3.6.1.32 Myosin ATPase 8e-09
 [PIRKW] blocked amino end 1e-07
 [PIRKW] nucleus 1e-06
 [PIRKW] citrulline 1e-07
 [PIRKW] tandem repeat 8e-09
 [PIRKW] heterodimer 3e-06
 [PIRKW] DNA repair 2e-06
 [PIRKW] heart 8e-09
 [PIRKW] endocytosis 3e-07
 [PIRKW] transmembrane protein 4e-10
 [PIRKW] zinc finger 3e-07
 [PIRKW] metal binding 3e-07
 [PIRKW] muscle contraction 8e-09
 [PIRKW] acetylated amino end 1e-06
 [PIRKW] actin binding 8e-09
 [PIRKW] microtubule binding 1e-06
 [PIRKW] cell division control 1e-06
 [PIRKW] ATP 8e-09
 [PIRKW] chromosomal protein 3e-06
 [PIRKW] thick filament 8e-09
 [PIRKW] phosphoprotein 1e-145
 [PIRKW] skeletal muscle 8e-09
 [PIRKW] calcium binding 1e-07
 [PIRKW] meiosis 2e-06
 [PIRKW] alternative splicing 7e-08
 [PIRKW] DNA condensation 3e-06
 [PIRKW] coiled coil 4e-10
 [PIRKW] P-loop 8e-09
 [PIRKW] heptad repeat 1e-07
 [PIRKW] methylated amino acid 8e-09
 [PIRKW] immunoglobulin receptor 2e-06
 [PIRKW] peripheral membrane protein 3e-07
 [PIRKW] cardiac muscle 8e-09
 [PIRKW] hydrolase 8e-09
 [PIRKW] muscle 7e-08
 [PIRKW] EF hand 1e-07
 [PIRKW] cytoskeleton 7e-08
 [PIRKW] hair 1e-07
 [PIRKW] smooth muscle 7e-08
 [PIRKW] calmodulin binding 3e-07
 [SUPFAM] conserved hypothetical P115 protein 2e-09
 [SUPFAM] myosin heavy chain 8e-09
 [SUPFAM] RAD50 protein 2e-06
 [SUPFAM] calmodulin repeat homology 1e-07
 [SUPFAM] myosin motor domain homology 8e-09
 [SUPFAM] alpha-actinin actin-binding domain homology 1e-06
 [SUPFAM] tropomyosin 7e-08
 [SUPFAM] protein-tyrosine kinase ret 3e-07
 [SUPFAM] plectin 1e-06
 [SUPFAM] trichohyalin 1e-07
 [SUPFAM] pleckstrin repeat homology 2e-06
 [SUPFAM] ribosomal protein S10 homology 1e-06
 [SUPFAM] protein kinase homology 3e-07
 [SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
 [SUPFAM] giantin 4e-06
 [SUPFAM] kinesin-related protein KLPA 1e-06
 [SUPFAM] kinesin motor domain homology 1e-06
 [SUPFAM] human early endosome antigen 1 3e-07
 [SUPFAM] M5 protein 2e-06
 [PROSITE] MYRISTYL 1
 [PROSITE] AMIDATION 1
 [PROSITE] CK2_PHOSPHO_SITE 6

```

SEQ      MNGTRNCTLVDPHPEDQAAGSVDIILRLTLQGLTGELEHIAQKAGRKYAMVSSHASG
SEG      .
PRD      cccccceeeeeeecccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .

SEQ      HSLASELVESHDGHEEIIKVYLKGRSGDKMIHEKNINQLKSEVQYIQEARNLCLKLREDI
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .

SEQ      SSKLDRLNGDSLHRQEIQVVLEKPNGFSQSPTALYSSPPEVDTCINEDVESLRKTVQDLL
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .CCCCCCCCCCCCCCCCCC

SEQ      AKLQEAQRHQSDCVAFEVTLSRYQREAEQSNVALQREEDRVEQKEAEVGELQRRLLGME
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCCCC.CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      TEHQALLAKVREGEVALEE LRSNNADCAEREAATLEKEVAGLREKIHHLDMLKSQQR
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CC.CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      KVRQMIEQLQNSKAVIQSKDATIQELKEKIAYLEAENLEMHDRMEHLIEKQISHGNFSTQ
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCC.CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      ARAKTENPGSIRISKPPSPKMPVIRVET
SEG      .xxxxxxxxxxxxxxxxxxxxxx..
PRD      hhccccccccceecccccccccceecccc
COILS    .

```

Prosites for DKFZphute1 19g22.3

PS00001	2->6	ASN_GLYCOSYLATION	PDOC00001
PS00001	356->360	ASN_GLYCOSYLATION	PDOC00001
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	171->174	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	157->161	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDOC00006
PS00008	355->361	MYRISTYL	PDOC00008
PS00009	46->50	AMIDATION	PDOC00009

(No Pfam data available for DKFZphut1 19g22.3)

DKFZphut1_19h17

group: intracellular transport and trafficking

DKFZphut1_19h17 encodes a novel 879 amino acid protein, with similarity to *N.crassa* osbP oxysterol-binding protein.

The novel protein contains a oxysterol-binding protein family signature. Mammalian oxysterol-binding protein (OSBP) is a protein binds a variety of oxysterols (oxygenated derivatives of cholesterol). OSBP seems to play a complex role in the regulation of sterol metabolism. OSBP is a cytosolic/Golgi receptor for oxysterols such as 25-hydroxycholesterol, and thus a potential target of siphingomyelin turnover and cholesterol mobilization at the plasma membrane and/or Golgi apparatus. Therefore, the new protein seems to be involved in oxysterol metabolism.

The new protein can find application in modulating the response of cells to oxysterols. The protein can be used as marker for the golgi system. The Protein might be used to direct drugs to the golgi system in response to oxidative stress.

strong similarity to *C.elegans* ZK1086.1 and oxysterol-binding proteins

complete cDNA, complete cds, few EST hits
similarity to proteins involved in steroid biosynthesis

Sequenced by AGOWA

Locus: unknown

Insert length: 3828 bp

Poly A stretch at pos. 3811, polyadenylation signal at pos. 3784

```

1  GCCCGCGCGC CCGGCCGGCC CGGAGCACCG AGCTCGCGGC ACGGTAGGAG
51  AAGCCCCCGA GCGCCACACG CATGAAGGAG GAGGCCTTCC TCCGGCGCCG
101 CTTCFCCCTG TGTCCACCTT CCTCCACCCC TCAGAAAGTC GACCCCGGGA
151 AGCTCACCCG GAACTTGCTC CTCAGCGGAG ACAATGAGCT CTACCCACTC
201 AGCCAGGAGA AGGACATGGA GCCCAACGGC CCGTCGCTGC CCAGGGATGA
251 AGGGCCCCCG ACCCAAGCT CTGCCACGAA GGTGCCACCG GCAGAGTACA
301 GGCTGTGCAA CGGGTCAGAC AAGGAATGTG TGTCCCCCAC CGCCAGGGTC
351 ACCAAGAAGG AGACTCTCAA GGCGCAGAAG GAGAACTACC GGCAGGAGAA
401 GAAGCGCGCC ACACGGCAGC TGCTCAGCGC TCTGACAGAC CCCAGCGTGG
451 TCATCATGGC TGACAGCCTG AAGATCCGCG GCACCTGAA GAGCTGGACC
501 AAGCTGTGGT GCGTGCTGAA GCCGGGGGTG CTGCTCATCT ACAAGACGCC
551 CAAGGTGGGC CAGTGGGTGG GCACGGTGCT GCTGCACTGC TCGAGCTCA
601 TCGAGCGGCC CTCCAAGAAG GACGGCTTCT GCTTCAAGCT CTTCCACCCG
651 CTGGATCAGT CCGTCTGGGC CGTGAAGGGC CCCAAGGTG AGAGCGTGGG
701 CTCCATCACA CAGCCCTGCG CCAGCAGCTA CCTGATCTTC AGGGCCCGCT
751 CCGAGTCAGA TGGTCGCTGC TGGCTGGACG CCCTGGAGCT GGCCCTGCGC
801 TGCTTAGCCG TACTGAGACT GGGCACCTGC AAGCCGGGCC GAGACGGGGA
851 GCCAGGGACC TCGCCAGACG CATCACCTC ATCGCTCTGT GGGCTGCCAG
901 CCTCAGCCAC TGTCCACCCA GACCAAGACC TGTTCCCACT GAACGGGTCT
951 TCCCTGGAGA ACGATGCATT CTCAGACAAG TCGGAGAGAG AGAACCCCTG
1001 GGAGTCAGAT ACCGAGACCC AGGACCATAG CCGGAAGACG GAGAGTGGCA
1051 GCGACCACTG AGAGACCCCT GGGGCCCGCG TCGGAGAGAG GACCACCTAT
1101 GTGGAGCAGG TCCAGGAGGA GCTGGGGGAG CTGGGCGAGG CGTCCCAGGT
1151 GGAGACAGTG TCAGAGGAGA ACAAGAGTCT GATGTGGACC CTGCTGAAGC
1201 AGCTACGGCC AGGCATGGAC CTGTCCCGCG TGGTGTACC CACGTTCTGA
1251 CTGGAGCCGC GCTCCTTCCT GAACAAGCTC TCCGACTACT ACTACCACGC
1301 AGACCTGCTC TCCAGGGCTG CGGTGGAGGA GGATGCCTAC AGCCGCATGA
1351 AGCTGGTGCT GCGGTGGTAC CTGTCTGGCT TCTACAAGAA GCCCAAGGGA
1401 ATCAAGAAGC CGTACAACCC CATCTGGGG GAGACCTTCC GCTGCTGTG
1451 GTTCCACCCG CAGACTGACA GCCGCACATT CTACATAGCA GAGCAGGTGT
1501 CCCACCACCC GCCCGTGTCT GCCTTCCACG TCAGCAACCG GAAGGACGGC
1551 TTCTGCATCA GTGGCAGCAT CACAGCCAAG TCCAGGTTTT ATGGGAATCT
1601 GCTGTGCGCG CTGCTGGACG GCAAAGCCAC GCTCACCTTC CTGAACCGAG
1651 CCGAGGATTA CACCCTTACC ATGCCCTACG CCCACTGCAA AGGAATCCTG
1701 TATGGCACGA TGACCTTGGG GCTGGGTGGG AAGGTCACCA TCGAGTGTGC
1751 GAAGAACAACT TTCCAGGCCG AGCTGGAATT CAAACTCAAG CCCTTCTTCG
1801 GGGGTAGCAC CAGCATCAAC CAGATCTCGG GAAAGATCAC GTCCGGAGAG
1851 GAAGTCTTGG CGAGCCTCAG TGGCCACTGG GACAGGGACG TGTATTATCA
1901 GGAGGAAGGG AGCGGAAGCA GTGCGCTTTT CTGACCCCGG AGCGGGGAGG
1951 TCCGCAGACA GAGGCTGAGG CAGCACACGG TGCCCGTGGG GGAGCAGACG
2001 GAGCTGGAGT CCGAGAGGCT CTGGCAGCAC GTACCCAGGG CCATCAGCAA
2051 GGGCGACCA CACAGGGCCA CACAGGAGAA GTTTCAGCTG GAGGAGGCAC
2101 AGCGGCAGCG GGCCCGTGAG CGGCAGGAGA GCCTCATGCC CTGGAAGCCG
2151 CAGCTGTTC CACTGGACCC CATCACCCAG GAGTGGCACT ACCGATACGA
2201 GGACCACAGC CCCTGGGACC CCCTGAAGGA CATCGCCAG TTTGAGCAAG
2251 ACGGATCCTT GCGGACCTTG CAGCAGGAGG CCGTGGCCCG CCAGACCAAC

```

```

2301 TTCTTGGGCA GCCCAGGGCC CAGGCACGAG AGGTCTGGCC CAGACCAGCG
2351 GCTTCGCAAG GCCAGCGACC AGCCCTCCGG CCACAGCCAG GCCACGGAGA
2401 GCAGCGGATC CACGCCTGAG TCCTGCCCGAG AGCTCTCAGA CGAGGAGCAG
2451 GATGGTGACT TTGTCCCTGG CGGTGAGAGC CCATGCCCTC GGTGCAGGAA
2501 GGAGGCGCGG CGGCTGCAGG CCCTGCACGA GGCCATCCTC TCCATCCGAG
2551 AGGCCAGACA GGAGCTGCAC AGGCACCTCT CGGCCATGCT GAGCTCCACG
2601 GCACGGGCAG CACAGGCACC GACCCAGGC CTCCTGCAGA GCGCCCGATC
2651 CTGGTTCTCT CTCTGCGTGT TCCTGGCGTG TCAGCTGTTC ATTAACCACA
2701 TCCTCAAATA GGAGCCCTGG GGGCAGAGCT CCTGGCCAGT CCCGAGCCCT
2751 CCCTCCCAAG CACCCAGCAC TTTAAGCCTG CTCATGGAG GCAGAGAGGC
2801 CCGCAAGCA CAGCCACTGT GACGGGAGT CCAGGCCAG GAGGGACCCG
2851 GGGCCACAAG GCGCTGCGGG CCCAGGTGTG CTGGGCCCT CTCAGGGGCA
2901 CTGGCCTCTC TGCAGGGCCT TCCGCCAGC GCTGGCCTTA ATGCTAAAGC
2951 CAAATGCAGC TTCTGCTGTG CGACGCACTC CTGGCCATCT TGCCGTGTCA
3001 CCCCTGTCC GGCCTCCACT TGCCATGGGG GATGGATGGA TTTAGGGTGG
3051 GAGGGCCTGT GGGGGCCCTG GACAGTCACA CCCAGCAGC AGTGAGTGGG
3101 CAGGTTTGGG GGAGCAGCCA GGGAGCCCG AGTGGCCAG GAGTCCCCCC
3151 ACACACAGAT GCATAGGCCT GCCTTCCGGA GACCTGTTC ACATTGCCGG
3201 GACCACCTGT GTGGGGCCAC TGGTGGGTGC CAGGGACAGG TTAGGGCCAC
3251 TCTGGGGAAG GCATTTTGGT TTTTATTCC ACGCTCTGCT GTTTGGATGG
3301 GAGCCCGACA GAGGCAGGTC CTGGAACAC CCCACCCCA CACCTGGACG
3351 CTCGCTCTGG TGGGGGCACA CGCAGGTGGA GGTGGTTGTG GGTGCAGGTG
3401 TGTGCAGGGG TGTGGGGGGC GCAGGGGTGT GGCTTAGCTG GCGCCGACC
3451 CAGGCCGGGG AGGCTCAAGT TCGCCACTTT ACTCAGACCG ATGCACAGTC
3501 TTCCCATTTT ACACTTTTTT AATAAACATA ATTGCAATAT TTTAGGTGGG
3551 CTGCGAGCTG CAGTCAGCCT TCACGTCTGG CCTCAGTCCC CGTGTCACTG
3601 CCGCTCTGCG TGTGCGTGTG CGCGTGTGTG AGCCTCTACA CATATATATA
3651 TGACACAGC CTAAACCAC ATCGTGGCGG TGCCGTCTGA GCTGTAGCGG
3701 GTGGCTTTGT TTCCAGTTTT TGTACCCGTG TCCTTGCTCT CCCTCCTCCC
3751 CCATCTGGGG ATGTGTCTGT GTTCCACACC TTGAAATAAA CAGACACATA
3801 CGTGTCTCT TAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

98315477:

The pleckstrin homology domain of oxysterol-binding protein recognises a determinant specific to Golgi membranes.

98146266:

A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

98146266:

A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

Peptide information for frame 3

ORF from 72 bp to 2708 bp; peptide length: 879

Category: strong similarity to known protein

```

1 MKEEAFLRRR FSLCPPSSSTP QKVDPRKLTR NLLSGDNEL YPLSPGKDME
51 PNGPSLRPRD GPPTPSSATK VPPAEYRLCN GSDKECVSPT ARVTKKETLK
101 AQKENYRQEK KRATRQLLSA LTDPSSVIMA DSLKIRGTLK SWTKLWCVLK
151 PGVLLIYKTP KVGQWVGTVL LHCCELIERP SKKDGFCFKL FHPLDQSVWA
201 VKGPKGESVG SITQPLPSSY LIFRAAESD GRCWLDALEL ALRCSLLRL
251 GTCKPGRDGE PGTSPDASPS SLCLPASAT VHPDQDLFPL NGSSLENDAL
301 SDKSERENPE ESDTETQDHS RKTESGSDQS ETPGAPVRRG TTYVEQVQEE
351 LGELGEASQV ETVSEENKSL MWTLKQLRP GMDLSRVVLP TFVLEPRSF
401 NKLSDDYYHA DLLSRAAVEE DAYSRMKLVL RWYLSGFYKK PKGIKKPYNP
451 ILGETFRCCN FHPQTDSTRTF YIAEQVSHHP PVSAFHVSNR KDGFCISGSI
501 TAKSRFYGNS LSALLDGKAT LTFLNRAEDY TLTMPYAHCK GILYGTMTLE
551 LGGKVTIECA KNNFQAQLEF KLPFFGGST SINQISKIT SGEEVLASLS
601 GHWRDRDVFIK EEGSGSSALF WTPSGEVRRO RLRQHTVPLE EQTELESERL

```

```

651 WQHVTRAISK GDQHRATQEK FALEEAQRQR ARERQESLMP WKPQLFHLDP
701 ITQEWHYRYE DHSPWDPLKD IAQFEQDGIL RTLQQEAVAR QTTFLGSPGP
751 RHERSGPDQR LRKASDQPSG HSQATESSGS TPESCPLESD EEQDGFVPG
801 GESPCPRCRK EARRLQALHE AILSIREAQQ ELHRHLSAML SSTARAAQAP
851 TPGLLQSPRS WFLLCVFLAC QLFINHILK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19h17, frame 3

TREMBL:CEZK1086_2 gene: "ZK1086.1"; *Caenorhabditis elegans* cosmid
ZK1086, N = 1, Score = 1495, P = 2.7e-153

PIR:S25324 hypothetical protein YKR003w - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 574, P = 8.5e-57

TREMBL:CEAF195_7 gene: "C32F10.1"; *Caenorhabditis elegans* cosmid
C32F10., N = 1, Score = 588, P = 8.6e-57

PIR:S46796 hypothetical protein YKR003w homolog YHR001w - yeast
(*Saccharomyces cerevisiae*), N = 1, Score = 585, P = 1.9e-56

TREMBL:NCOSBP_1 gene: "osbp"; product: "oxysterol-binding protein";
N.crassa mRNA for putative oxysterol-binding protein, N = 1, Score =
571, P = 7e-55

TREMBL:AB017026_1 product: "oxysterol-binding protein"; *Mus musculus*
mRNA for oxysterol-binding protein, complete cds., N = 2, Score = 328,
P = 3e-35

>TREMBL:CEZK1086_2 gene: "ZK1086.1"; *Caenorhabditis elegans* cosmid ZK1086
Length = 751

HSPs:

Score = 1495 (224.3 bits), Expect = 2.7e-153, P = 2.7e-153
Identities = 327/663 (49%), Positives = 430/663 (64%)

```

Query:  129 MADSLKIRGTLKSWTKLWCVLKPGVLLIYKTPKV--GQWVGTVLLHCCELIERPSKKDGF 186
        MAD+LKIRG LK W + +CVLKPGL++YK K G WVGTVLL+ CELIERPSKKDGF
Sbjct:   1  MADTLKIRGALKRWNRYYCVLKPGLLILYKHKKADRGDWVGTVLLNHCCELIERPSKKDGF 60

Query:  187 CFKLFHPLDQSVWAVKGPKGESVGSIT-QPLPSSYLIFRAASESDGRCWLDALALRCS 245
        CFKLFHFP+D S+W +GP G+S GS T PL +S+LI RA S+ GRCW+DALEL+ +C+
Sbjct:   61 CFKLFHPMDMSIWGNRGLGQSGSFTLNPLNTSFLICRAPSDQAGRCWMDALELSFKCT 120

Query:  246 SLLRLGTCKPGRDGEFGTSPDASPSSLCGLPASATVHPDQDLFPLNGSSLENDAFSDK-S 304
        LL+ T D + G D+S + G + + D D G A S+ +
Sbjct:  121 GLKKK-TMNE-LDDKNG---DSSMND--GQRDESMSRSDS-----GDDTRELAVSETDA 168

Query:  305 ERENPEESDTEQDHSRKTESGSDQSETPGAPVRRGTT---YVEQVQEELGELGEASQVE 361
        E+ E D + +DH E G SET +R T ++ +E G G S E
Sbjct:  169 EKHFQEIDDVQDEDH---EDGK-MSETSOT-IREAFTESAWIPSPKEVFGPDG--SLTE 220

Query:  362 TVSEENKSLMWTLKQLRPGMDLSRVVLPFTFVLEPRSFLNKLSDYYYHADLLSRAAVEED 421
        V EENKSL+WTLLKQ+RPGMDLS+VVLPTF+LEPRSFL KL+DYYYHADL+S A E D
Sbjct:  221 EVGEENKSLIWTLLKQIRPGMDLSKVVLPTFILEPRSFLEKLADYYYHADLISEAVAEPD 280

Query:  422 AYSRMKLVLRWYLSGFYKPKGIKKFPYNPILGETFRCCWFHPQDTSRTFYIAEQVSHHPP 481
        + R+ V +++LSGFYKPKG+KKFPYNPILGETFRC W HP S TFY+AEQVSHHPP
Sbjct:  281 PFQRIVKVTKFLLSGFYKPKGLKKFPYNPILGETFRCKWEHPD-GSTTFYMAEQVSHHPP 339

Query:  482 VSAFHVSNRKGDFCISGSITAKSRFYGNLSALLDGKATLTFLNRAEDYTLTMPYAHCKG 541
        VS+ ++NRK GF ISG+I AKS++YGNLSLA+L GK LT LN E Y + +PYA+CKG
Sbjct:  340 VSSLFITNRKAGFNISGTLAKSKYYGNLSAILAGKLRLLTLLNLGETYIIVNLPYANCKG 399

Query:  542 ILYGTMTELEGGKVITCAKNNFQAQLEFKLPFFGGGSTSINQISGKITSGEVVLASLSG 601
        I+ GTMT+ELGG+V IEC K ++ L+FKLKP GG+ NQI G I G + LAS+ G
Sbjct:  400 IMIGTMTMELGGEVNIIECKTGYRTTDLFKLPMLGGA--YNQIEGSIKYGSDRLASIEG 457

Query:  602 HWRDVFIIKEEGSGSSALFWTPSGEVRRQRLRQHTVPLEEQTELESERLWQHVTRAISKG 661
        WD + IK G W P+ EV + RL ++ + ++EQ E ES +LW+HVT AIS
Sbjct:  458 AWDGVIRIK--GPDGKKELWNPTPEVIKTRLPRIEINMDEQGEWESAKLWRHVTEAISNE 515

Query:  662 DQHRATQEKFALEEAQRQRARERQESLMPWKPQLFHLDPITQEWHYRYEDHSPWDPLKDI 721
        DQ++AT+EK ALE QR RA+ S +P + + F ++ Y + D+ PWD DI
Sbjct:  516 DQYKATEEKTALENDQARAK----SGIPHETKFFKKQH-GDDYVYIHADYRPWDNNNDI 570

```

```

Query:      722 AQFEQDGILRTLQQEAVAR--QTTFLGSGPGRHERSGPDQRLRKASDQPSGHSQATESSG 779
             Q E + +++T+ + + + LGS E S D+ + +P + + +
Sbjct:      571 QQIENNYVVKTISRHSKRKTGNSEQLGSNDNTS-EASESDEEVI---EPKIKKKKEIVPAK 625

Query:      780 STPESCPELSDE 791
             S P + PE++DE
Sbjct:      626 SKPIT-PEVADE 636

```

Pedant information for DKFZphutel_19h17, frame 3

Report for DKFZphute1 19h17.3

```

[LENGTH]      879
[MW]           98616.79
[pI]           7.29
[HOMOL]        TREMBL:CEZK1086_2 gene: "ZK1086.1"; Caenorhabditis elegans cosmid ZK1086 1e-157

[FUNCAT]       01.06.16 lipid and fatty-acid binding           [S. cerevisiae, YHR001w] 3e-55
[FUNCAT]       01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR001w]
3e-55
[FUNCAT]       30.03 organization of cytoplasm                 [S. cerevisiae, YPL145c] 3e-23
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YPL145c]
3e-23
[FUNCAT]       04.05.01.07 chromatin modification             [S. cerevisiae, YAR044w] 5e-20
[BLOCKS]       BL00168F
[BLOCKS]       BL01013D Oxysterol-binding protein family proteins
[BLOCKS]       BL01013C Oxysterol-binding protein family proteins
[BLOCKS]       BL01013B Oxysterol-binding protein family proteins
[BLOCKS]       BL01013A Oxysterol-binding protein family proteins
[PIRKW]        transmembrane protein 1e-19
[SUPFAM]       pleckstrin repeat homology 8e-18
[SUPFAM]       ankyrin repeat homology 1e-19
[SUPFAM]       unassigned ankyrin repeat proteins 1e-19
[PROSITE]      MYRISTYL 12
[PROSITE]      CAMP_PHOSPHO_SITE 6
[PROSITE]      OSBP 1
[PROSITE]      CK2_PHOSPHO_SITE 21
[PROSITE]      PROKAR_LIPOPROTEIN 1
[PROSITE]      TYR_PHOSPHO_SITE 2
[PROSITE]      PKC_PHOSPHO_SITE 20
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         PH (pleckstrin homology) domain
[KW]           TRANSMEMBRANE 1
[KW]           LOW COMPLEXITY 2.96 %
[KW]           COILED COIL 3.53 %

```

[illegible]

```

SEQ      GPPTPSSATKVPAPAEYRLCNGSDKECVSPRTARVTKKETLKAQKENYRQEKKRATRQLLSA
SEG      .....
PRD      cccccccccccccccceccccccceccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....cccccccccccccccccccccccccccccccccc
MEM

```

```
SEQ      LTDPSPVIMADSLKIRGTLKSWTKLWCVLKPGLLIYKTPKVQWVGTVLLHCCCELIERP
SEG      .....
PRD      hccccceeeccccccccccccccceeeeeeccccceeeeccccccccccc
COILS    CCC.....
MEM
```

[illegible][illegible]

SEQ SDKSERENPEESDTETQDHSRKTESGSDQSETPGAPVRRGTTYVEQVQEELGELGEASQV

Prosites for DKFZphute1 19h17.3

473

PS00005	301->304	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	455->458	PKC_PHOSPHO_SITE	PDOC00005
PS00005	488->491	PKC_PHOSPHO_SITE	PDOC00005
PS00005	501->504	PKC_PHOSPHO_SITE	PDOC00005
PS00005	586->589	PKC_PHOSPHO_SITE	PDOC00005
PS00005	647->650	PKC_PHOSPHO_SITE	PDOC00005
PS00005	824->827	PKC_PHOSPHO_SITE	PDOC00005
PS00005	843->846	PKC_PHOSPHO_SITE	PDOC00005
PS00005	857->860	PKC_PHOSPHO_SITE	PDOC00005
PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	263->267	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	362->366	CK2_PHOSPHO_SITE	PDOC00006
PS00006	590->594	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00006	659->663	CK2_PHOSPHO_SITE	PDOC00006
PS00006	713->717	CK2_PHOSPHO_SITE	PDOC00006
PS00006	755->759	CK2_PHOSPHO_SITE	PDOC00006
PS00006	780->784	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00006	789->793	CK2_PHOSPHO_SITE	PDOC00006
PS00006	824->828	CK2_PHOSPHO_SITE	PDOC00006
PS00007	402->409	TYR_PHOSPHO_SITE	PDOC00007
PS00007	415->424	TYR_PHOSPHO_SITE	PDOC00007
PS00008	137->143	MYRISTYL	PDOC00008
PS00008	163->169	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	326->332	MYRISTYL	PDOC00008
PS00008	381->387	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	508->514	MYRISTYL	PDOC00008
PS00008	541->547	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	613->619	MYRISTYL	PDOC00008
PS00008	728->734	MYRISTYL	PDOC00008
PS00013	860->871	PROKAR_LIPOPROTEIN	PDOC00013
PS01013	474->485	OSBP	PDOC00774

Pfam for DKFZphut1_19h17.3

HMM_NAME	PH (pleckstrin homology) domain		
HMM	*dvIREGWMYKWgswrkstgnWqrRWFvLrndpnrLiYYkddkdekPrYM		
	+VI+ +++++G + W + W+VL++ ++L+ YK + + + ++		
Query	126	VVIMADSLKIRGTLKS----WTKLWCVLKP--GVLLIYKTP-KVGQWVG	167
HMM	l1ldlcWrMidVEidWmmdndHCFiIWtrq.....		
	L+C+ +I+ ++ ++ +CF+++ +		
Query	168	TVLLHCCELIERPSKKD--GFCFKLFHPLDQSVWAVKGPKGESVGSITQ	214
HMMrtYYFQAeNeEEMmeWsaIrRaIw*		
	+ ++F+A++E++ + W++A++ A++		
Query	215	PLPSSYLIFRAASESDGRCWLDALALALR	243

DKFZphutel_19j11

group: uterus derived

DKFZphutel_19j11 encodes a novel 708 amino acid protein with C-terminal similarity to several known proteins, such as human KIAA0231 or murine ras binding protein Sur8.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

Strong similarity to KIAA0231, similarity to ras binding protein Sur8

EST AA854189 extends the sequence (294 Bp), with this sequence complete cDNA,

Sequenced by AGOWA

Locus: unknown

Insert length: 2343 bp

Poly A stretch at pos. 2323, polyadenylation signal at pos. 2295

```
1 GCTCCTGCTA ACCCCATCAC TGTGGAAATG AAAGGCCTGA AGACAGATT
51 GGACCTTCAG CAGTACAGCT TTATAAATCA GATGTGTTAT GAGCGAGCCC
101 TCCACTGGTA TGCCAAGTAT TTCCCTTACC TTGTCCTCAT CCATACCCCTG
151 GTCTTTATGC TCTGCAGTAA CTTTGGGTTT AAATTCCTCG GTTCCAGCTC
201 CAAAATAGAA CATTTCATCT CCATTCTGGG GAAGTGTTT GACTCTCCTT
251 GGACCACACG GGCTTTATCT GAAGTGTCTG GGGAGGACTC AGAAGAAAAG
301 GACAACAGGA AGAACAACAT GAACAGGTCC AACACCATCC AATCTGGTCC
351 AGAAGGCAGC CTGGTCAACT CTCAGTCTTT AAAGTCCATT CCTGAGAAGT
401 TTGTAGTTGA TAAATCCACT GCAGGGGCTC TGGATAAAAA GGAAGGTGAG
451 CAGGCTAAGG CCTTATTTGA GAAGGTGAAG AAGTTCAGGC TGCATGTGGA
501 AGAAGGTGAT ATTCTATATG CCATGTATGT TCGCCAGACT GTACTTAAAG
551 TTATCAAAAT CCTAATCATC ATTGCATATA ATAGTGCTCT GGTTCCTCAAG
601 GTCCAGTTTA CAGTGGACTG TAATGTGGAC ATTCAGGACA TGAAGTGATA
651 TAAAAACTTT TCTTGCAATC ATACCATGGC ACACCTGTTC TCAAACTGT
701 CCTTTTGCTA TCTGTGCTTT GTTAGTATCT ATGGATTGAC GTGCCTTTAT
751 ACCTTATACT GGCTGTTCTA CCGTTCCTA CGGGAATATT CCTTTGAGTA
801 TGTCCTGCTG GAGACTGGAA TTGATGATAT TCCAGATGTG AAAAATGACT
851 TTGCTTTTAT GCTTCATATG ATAGATCAGT ATGACCTCTC CTATTCCAAG
901 AGATTTGAGC TGTTCTGCTC TGAAGTCAGT GAAACAAAT TAAAGCAGCT
951 GAACCTTAAT AACGAATGGA CTCCTGATAA ACTGAGGCAG AAGCTACAGA
1001 CAATGCCCCA TAATCGACTG GAATTGCCCT TTATCATGCT CTCTGGCCTT
1051 CCAGACACTG TTTTGAAT CACAGAGTTG CAATCTCTAA AACTTGAAAT
1101 CATTAAGAAC GTAATGATAC CAGCCACCAT TGCACAGCTA GACAATCTTC
1151 AAGAGCTCTC TCTGCACCAG TGTCTGTCTA AAATCCACAG TCGCGCGCTC
1201 TCTTTCTGTA AGGAAAACCT CAAGGTCTTG AGCGTCAAGT TTGATGACAT
1251 GAGGGAACCT CCCCCTGGA TGTATGGGCT CCGAAATCTG GAAGAGCTGT
1301 ACCTAGTTGG CTCTCTAAGT CATGATATTT CCAGAAATGT CACCTTGAG
1351 TCTCTCGGGG ATCTCAAAAG CCTTAAATTT CTCTCTATCA AAAGCAACGT
1401 TTCCAAATAT CCTCAGGCAG TGGTTGATGT TTCCAGCCAT CTCCAGAAGA
1451 TGTGCATACA TAATGATGGC ACCAAGCTGG TGATGCTCAA CAACCTAAAG
1501 AAGATGACCA ATCTGACAGA GCTGGAGCTG GTCCACTGTG ACCTGGAGCG
1551 TATTCCTCAT GCTGTGTTCA GCCTACTCAG CCTCCAGGAA TTGGACCTGA
1601 AGGAAAACAA TCTGAAATCT ATAGAAGAAA TCGTTAGCTT TCAGCACTTA
1651 AGAAAGTTGA CAGTGCTAAA ACTGTGGCAT AACAGCATCA CCTACATCCC
1701 AGAGCATATA AAGAAATCA CCAGCCTGGA ACGCCTGTCC TTTAGTCACA
1751 ATAAAAATAGA GGTGCTGCCT TCCCACCTCT TCCTATGCAA CAAGATCCGA
1801 TACTTGGACT TATCGTACAA TGACATTGCA TTTATCCCCC CTGAAATTGG
1851 AGTTCTACAA AGTTTACAGT ATTTTCCAT CACATGTAAC AAAGTGGAAG
1901 GCCTTCCAGA TGAACCTTAC TTCTGCAAGA AACTTAAAC TCTGAAGATT
1951 GGAATAAACA GCCTATCTGT ACTTTCACCG AAAATTGGAA ATTTGCTATT
2001 TCTTTCCTAC TTAGATGTAA AAGGTAATCA CTTTGAAATC CTCCTCCTG
2051 AACTGGGTGA CTGTCGGGCT CTGAAGCGAG CTGGTTTAGT TGTAAGAGAT
2101 GCTCTGTTTG AAATCTGTCC TTCTGACGTC CGGGAGCAAA TGAAACAGA
2151 ATAACTTATT TTTGTTTAAA GTTTGACTGA AACACGCTTC TACCAATAC
2201 AGTATAAATA ATTAGGTAGT CTTAATGCCT TTCCTATTTT TTTTTCCTTT
2251 TCACACAAAA TGTACACAAA GATCGCGTAA GGAGTATGTA TTTTAAATAA
2301 AAATTTAATT GTATTTTTTC AATATTAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

96421675:
 Characterization of densin-180, a new brain-specific synaptic protein
 of the
 O-sialoglycoprotein family.

98337190:
 SUR-8, a conserved Ras-binding protein with leucine-rich
 repeats, positively regulates Ras-mediated signaling in *C.*
elegans.

Peptide information for frame 1

ORF from 28 bp to 2151 bp; peptide length: 708
 Category: similarity to known protein
 Classification: Cell signaling/communication

```

1 MKGLKTDLDL QQYSFINQMC YERALHWHYAK YFPYLVLHIT LVFMLCSNFW
51 FKFPGSSSKI EHFISILGKC FDSPWTTRAL SEVSGEDSEE KDNRRKNMNR
101 SNTIQSGPEG SLVNSQSLKS IPEKFVVDKS TAGALDKKEG EQAKALFEKV
151 KKFRLHVEEG DILYAMYVRQ TVLKVIKFLI IAYNSALVS KVQFTVDCNV
201 DIQDMTGYNK FSCNHTMAHL FSKLSFCYLC FVSIYGLTCL YTLYWLFYRS
251 LREYSFEYVR QETGIDDIPD VKNDFAFMLH MIDQYDPLYS KRFAVFLSEV
301 SENKQLQLNL NNEWTPDKLR OKLQTNANHR LELPLIMLSG LPDVTVEITE
351 LQSLKLEIHK NVMIPATIAQ LDNLQELSLH QCSVKIHSAA LSFLKENLKV
401 LSVKFDDMRE LPPWMYGLRN LEELYLVGSL SHDISRNVTL ESLRDLKSLK
451 ILSIKSNVSK IPQAVVDVSS HLQKMCIHND GTKLVMNLNL KMTNLTELE
501 LVHCDLERIP HAVFSLLSLQ ELDLKENNLK SIEEIVSFQH LRKLTVLKLW
551 HNSITYIPEH IKKLTSLERL SFSHNKIEVL PSHLFLCNKI RYLDLSYNDI
601 RFIPPEIGVL QSLQYFSITC NKVESLPDEL YFCKKLKTLK IGKNSLSVLS
651 PKIGNLLFLS YLDVKGNHFE ILPPELGDCR ALKRAGLVVE DALFETLPSD
701 VREQMKTE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19j11, frame 1

TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene,
 partial cds., N = 1, Score = 1408, P = 4.5e-144

TREMBL:AF054827_1 gene: "soc-2"; product: "leucine-rich repeat protein
 SOC-2"; *Caenorhabditis elegans* leucine-rich repeat protein SOC-2
 (soc-2) mRNA, complete cds., N = 1, Score = 304, P = 5.7e-24

TREMBL:RNU66707_1 product: "densin-180"; *Rattus norvegicus* densin-180
 mRNA, complete cds., N = 1, Score = 311, P = 7.4e-24

TREMBL:AF068921_1 product: "Ras-binding protein SUR-8"; *Mus musculus*
 Ras-binding protein SUR-8 mRNA, complete cds., N = 1, Score = 302, P =
 1.1e-23

>TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial
 cds.

Length = 476

HSPs:

Score = 1408 (211.3 bits), Expect = 4.5e-144, P = 4.5e-144
 Identities = 265/471 (56%), Positives = 361/471 (76%)

Query: 237 LTCLYTLWLFYRSREYSFEYVRQETGIDDIPDVKNDFAFMLHMIDQYDPLYSKRFAVF 296
 LT Y+L+W+ SL++YSFE +R+++ DIPDVKNDFAF+LH+ DQYDPLYSKRF++F
 Sbjct: 1 LTSSYSYLWMLRSSLKQYSFEALREKSNYSIDIPDVKNDFAFILHLADQYDPLYSKRFSIF 60

Query: 297 LSEVSENKQLQLNLNNEWTPDKLRQKLQTNANRLELPLIMLSGLPDTVFEITELQSLKL 356
 LSEVSENKQLKQ+NLNNEWTPDKLRQKLQTNANRLELPLIMLSGLPDTVFEITELQSLKL 356
 Sbjct: 61 LSEVSENKQLQINLNNEWTPDKLRQKLQTNANRLELPLIMLSGLPDTVFEITELQSLKL 120

Query: 357 EIIKNVMPATIAQLDNLQELSLHQCSVKIHSAAFLKENLKVSVKFDMDRELPPWMY 416
 E+I V +P+ ++QL NL+EL ++ S+ + AL+FL+ENLK+L +KF +M ++P W++
 Sbjct: 121 ELIPEVKLPASVSQLVNLKELRVYHSSLVVDHPALAFLEENLKILRLKFTMGKIPRWVF 180

Query: 417 GLRNLEELYLVGSLSHDISRNVTLSELRDLKSLKILSIKSNVSKIPQAVVDVSSHLQKMC 476
 L+NL+ELYL G + + + LE +DLK+L+ L +KS++S+IPQ V D+ LQK+
 Sbjct: 181 HLKNLKELYLSGCVLPEQLSTMQLLEGFQDLKLNRLTYLKSLSRIPQVVTDLPLSLQKLS 240

Query: 477 IHNDGTKLVMLNNLKKMTNLTELELVHCDLERIPHAVFSLLSLQELDLKENNLKSIEEIV 536
 + N+G+KLV+LNNLKKM NL LEL+ CDLERIPH++FSL +L ELDL+ENNLK++EEI+
 Sbjct: 241 LDNEGSKLVVLNNLKKMVNLKSLELISCDLERIPHSIFSLNNLHEDLDRENNLKTVEEII 300

Query: 537 SFQHLRKLTVLKLWHNSITYIPEHIKKLTSLERLSFSHNKIEVLPSHLFLCNKIRYLDLS 596
 SFQHL+ L+ LKLWHN+I YIP I L++LE+LS HN IE LP LFLC K+ YLDLS
 Sbjct: 301 SFQHLQNLSCCLKLWHNNIAYIPAQIGALSNLQELSDHNNIENLPLQLFLCTKLHYLDLS 360

Query: 597 YNDIRFIPPEIGVLSQYFSITCNKVESLPDELYFCKKLKTLKIGKNSLSVLSPKIGNL 656
 YN + FIP EI L +LQYF++T N +E LPD L+ CKKL+ L +GKNSL LSP +G L
 Sbjct: 361 YNHLTFIPPEIIQLSNLQYFAVTNNIEMLPDGLFQCKKLQCLLLGKNSLMNLSPHVGEL 420

Query: 657 LFLSYLDVKGNHFEILPPELGDCRALKRAGLVVEDALFETLPDVRQMKMT 707
 L++L++ GN+ E LPPEL C++LKR L+VE+ L TLP V E+++T
 Sbjct: 421 SNLTHLELIGNYLETLPELEGCSLKRNLIVEENLLNTLPLPVTERTLQT 471

Pedant information for DKFZphutel_19j11, frame 1

Report for DKFZphutel_19j11.1

[LENGTH] 708
 [MW] 81812.82
 [pI] 7.55
 [HOMOL] TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial cds.
 1e-149
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae,
 YJL005w] 3e-17
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL193c] 3e-09
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
 palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 3e-09
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YAL021c] 9e-08
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YAL021c]
 9e-08
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YAL021c]
 9e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR353c] 3e-07
 [BLOCKS] BL00868F
 [BLOCKS] BL00985B Spermadhesins family proteins
 [EC] 3.4.17.3 Lysine carboxypeptidase 1e-08
 [EC] 4.6.1.1 Adenylate cyclase 3e-18
 [PIRKW] blocked amino end 1e-10
 [PIRKW] phosphotransferase 1e-09
 [PIRKW] nucleus 6e-08
 [PIRKW] duplication 3e-18
 [PIRKW] platelet 1e-10
 [PIRKW] tandem repeat 7e-16
 [PIRKW] keratan sulfate 7e-07
 [PIRKW] metallo-carboxypeptidase 1e-08
 [PIRKW] transmembrane protein 1e-10
 [PIRKW] serine/threonine-specific protein kinase 1e-09
 [PIRKW] autophosphorylation 1e-09
 [PIRKW] cartilage 7e-07
 [PIRKW] connective tissue 7e-07
 [PIRKW] magnesium 1e-09
 [PIRKW] cAMP biosynthesis 3e-18
 [PIRKW] ATP 1e-09
 [PIRKW] receptor 1e-09
 [PIRKW] leucine zipper 3e-13
 [PIRKW] glycoprotein 5e-12
 [PIRKW] extracellular matrix 7e-07
 [PIRKW] chondroitin sulfate proteoglycan 7e-07
 [PIRKW] cell adhesion 1e-08
 [PIRKW] hydrolase 1e-08
 [PIRKW] sulfoprotein 7e-07
 [PIRKW] membrane protein 1e-08
 [PIRKW] phosphorus-oxygen lyase 3e-18

```

[PIRKW] collagen binding 7e-07
[SUPFAM] leucine-rich alpha-2-glycoprotein repeat homology 3e-21
[SUPFAM] chaptain 1e-08
[SUPFAM] gelsolin repeat homology 3e-21
[SUPFAM] protein kinase homology 1e-09
[SUPFAM] protein kinase Xa21 1e-09
[SUPFAM] fibromodulin 4e-12
[SUPFAM] yeast adenylate cyclase catalytic domain homology 3e-18
[SUPFAM] yeast adenylate cyclase 3e-18
[KW] TRANSMEMBRANE 3
[KW] LOW COMPLEXITY 1.41 %

```

```
SEQ      MKGLKTDLDLQOYSFINQMCIERALHWYAKYPFYLVLHTLVFMLCNSFWFKPGSSSKI
SEG      .
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhcccceeeccccccee
MEM      .mmmmmmmmmmmmmmmmmmmmmm
```

```
SEQ      EHFISILGKCFDSPWTTTRALSEVSGEDSEEDNRKNNMNRSTIQSGPEGLVNSQSLKS
SEG      .....
PRD      eeeeeeeccccccccceeeccccccccccccccccccccccccccccccccccccccccc
MEM
```

```
SEQ      IPEK FVVDKSTAGALDKKEGEQAKALFEKVVKFRLHVEEGDILYAMYVRQTVLKVIKFLLI
SEG      .....
PRD      cccceeecccccccchhhhhhhhhhhhhhhhhhhhhc cccceeehhhhhhhhhhhhhhhhhhh
MEM      MMMMMMMMMM
```

```
SEQ      IIAYNSALVSKVQFTVDCNVDIQDMTGYKNFSCNHTMAHLSFKSLFCYLCFVSIIYGLTCL
SEG
PRD      hhhhccchhhheeeeeeccccccccccccccccccccchhhhhhhhheeeeeeeeeecceeee
MEM      mmmmmmmm                                     mmmmmmmmmmmmmmmmmmm
```

```
SEQ      YTLYLWLFYRSLREYSFEYVRQETGIDDIPDVKNDFAFMLHMDIQDYDPLYSKRFAVFLSEV  
SEG      .....  
PRD      hhhhhhhhhhhhhhhhhhhhhhcccccccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhh  
MEM
```

```
SEQ  SENK LKQLN LNN E W T P D K L R Q L Q T N A H R L E P L I M L S G L P D T V F E I T E L Q S L K L E I I K
SEG  . . . . . X X X X X X X X X X . . . . .
PRD  h h h h h h h h h c c c c c c c h h h h h h h h h h h h h h h h h c c c c c c c h h h h h h h h h h h h h h h h h
MEM
```

```
SEQ      NVMIPIATIAQLDNLQELSHQCSVKIHSAAFLKENLKVLSVKFDDMRELPPWMYGLRN
SEG      .
PRD      hccccccchhhhhhhhhhhhhccccccccccccchhhhhhhhhhhccccccccccchhh
MEM
```

```
SEQ      LEELYLVGSLSHDISRNVTLESRLDKSLKILSIKSNVSKIPQAVVDVSSH LQKMCIHND
SEG      .....
PRD      hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM
```

```
SEQ      GTKLVMLNNLKKMTNLTELELVHCDLERIPHAVFSLLSIQELDLKENNLSIEEIVSFQH
SEG      .....
PRD      ceeeeccccccchhhhhhhhhccccccccchhhhhhhhhhhccccccccccccch
MEM
```

```
SEQ      LRKLTVLKLVHNSITYIPEHIKKLTSLERLSFSHNKIEVLPSHLFLCNKIRYLDLSYNDI
SEG      .
PRD      hhhhhhhccccceeeccccchhhhhheeeccccceeeccccchhhhhhhhhcccc
MEM
```

```
SEQ      RFIPPEIGVLQSLQYFSITCNKVESLPDELYFCKKLKTLKIGKNSLSVLSPKIGNLLFLS
SEG      .....
PRD      cccccccchhhhhhhhhhhccccccccccccchhhhhcccccccceeeccccccchhhh
MEM
```

```
SEQ      YLDVKGHNHFEILPPELGDCRALKRAGLVVEDALFETLPDVRQMKTE
SEG      .....
PRD      hhhccccccccccccchhhhhhhhhheeecccccccccccccccccc
MEM
```

(No Prosite data available for DKFZphut1 19j11.1)

(No Pfam data available for DKFZphut1 19j11.1)

DKFZphut1_li2

group: transcription factor

DKFZphut1_li2 encodes a novel 594 amino acid protein similar to signal transducing proteins.

The protein contains 2 WD-40 repeats, which is typical for the beta-transducin subunit of G-proteins. In addition, the protein contains a C3HC4 zinc finger and a leucine zipper. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription.

The new protein can find application in modulating/blocking gene expression of genes controlled by this molecule.

similarity to Dictostelium myosin heavy chain kinase

complete cDNA, complete cds, EST hits
 [PFAM] Zinc finger, C3HC4 type (RING finger)
 [PFAM] WD domain, G-beta repeats
 [SCOP] dltbgc_2.46.3.1.1 beta1-subunit of the
 signal-transducing G protei 3e-07

Sequenced by BMF2

Locus: /map="16p13.3"

Insert length: 3584 bp

Poly A stretch at pos. 3555, polyadenylation signal at pos. 3537

```

1 GGGCGGGAGG TGCTTCCCAA GGACCGTAGA TGCCTCTCTA GAGCATGAGC
51 TCAGGCAAGA GTGCCCCGCTA CAACCGCTTC TCCGGGGGGC CCAGCAATCT
101 TCCACCCCCA GACGTCACCA CAGGGACCAG AATGGAAACG ACCTTCGGAC
151 CCGCCTTTTC AGCCGTCACC ACCATCACA AAGCTGACGG GACCAGCACC
201 TACAAGCAGC ACTGCAGGAC AGCATGCCCC CCATCAGCAC TCCCGGCCGC
251 TCCGACTCCG CCATCTCTGT CCGCTCCCTG CACTCAGAGT CCAGCATGTC
301 TCTGCCGTCC ACATTCTCAC TGCCCGAGGA GGAGGAGGAG CCGGAGCCAC
351 TGGTGTTTGC GGAGCAGCCC TCGGTGAAGC TGTGCTGTCA GCTCTGCTGC
401 AGCGTCTTCA AAGACCCCGT GATCACCACG TGTGGGCACA CGTCTGTAG
451 GAGATGCGCC TTGAAGTCAG AGAAGTGTC CGTGGACAAC GTCAAACTGA
501 CCGTGTGTGT GAACAACATC GCGGTGCGCG AGCAGATCGG GGAGCTCTTC
551 ATCCACTGCC GGCACGGCTG CCGGGTAGCG GGCAGCGGGA AGCCCCCAT
601 CTTTGAGGTG GACCCCGAG GGTGCCCTTC CACCATCAAG CTCAGCGCCC
651 GGAAGGACCA CGAGGGCAGC TGTGACTACA GGCTGTGCG GTGTCCCAAC
701 AACCCAGCTG GCGCCCGCT GCTCAGGATG AACCTGGAGG CCCACCTCAA
751 GGAGTGCAGC CACATCAAAAT GCCCCACTC CAAGTACGGG TGCACGTTCA
801 TCGGGAACCA GGACACTTAC GAGACCCACC TGGAGACTTG CCGCTTCGAG
851 GGCCTGAAGG AGTTTCTGCA GCAGACGGAT GACCGCTTCC ACGAGATGCA
901 CGTGGCTCTG GCCCAGAAGG ACCAGGAGAT CGCCTTCCTG CGCTCCATGC
951 TGGGAAAGCT CTCGGAGAAG ATCGACCAGC TAGAGAAGAG CCTGGAGCTC
1001 AAGTTTGACG TCCTGGACGA AAACCAGAGC AAGCTCAGCG AGGACCTCAT
1051 GGAGTTCCGG CGGGACGCAT CCATGTTAAA TGACGAGCTG TCCACATCA
1101 ACGCGCGGCT GAACATGGGC ATCCTAGGCT CCTACGACCC TCAGCAGATC
1151 TTCAAGTGCA AAGGGACCTT TGTGGGCCAC CAGGGCCCTG TGTGGTGTCT
1201 CTGCGTCTAC TCCATGGGTG ACCTGCTCTT CAGTGGCTCC TCTGACAAGA
1251 CCATCAAGGT GTGGGACACA TGTACCACCT ACAAGTGTC AAGACACTG
1301 GAGGGCCATG ATGGCATCGT GCTGGCTCTC TGCATCCAGG GGTGCAAACT
1351 CTACAGCGGC TCTGCAGACT GCACCATCAT TGTGTGGGAC ATCCAGAACC
1401 TGCAGAAGGT GAACACCATC CCGGCCCATG ACAACCCGGT GTGCACGCTG
1451 GTCTCCTCAC ACAACGTGCT CTTACAGCGC TCCCTGAAGG CCATCAAGGT
1501 CTGGGACATC GTGGGCACTG AGCTGAAGTT GAAGAAGGAG CTCACAGGCC
1551 TCAACCACTG GGTGCGGGCC CTGGTGGCTG CCCAGAGCTA CCTGTACAGC
1601 GGCTCCTACC AGACAATCAA GATCTGGGAC ATCCGAACCC TTGACTGCAT
1651 CCACGTCCTG CAGACGTCGT GTGGCAGCGT CTAATCCATT GCTGTGACAA
1701 ATCACCACAT TGTCTGTGGC ACCTACGAGA ACCTCATCCA CGTGTGGGAC
1751 ATTGAGTCCA AGGAGCAGGT GCGGACCCTC ACGGGCCACG TGGGCACCGT
1801 GTATGCCCTG GCGGTCACTC CGAGCCGAGA CCAGACCAAA GTCTTCAGTG
1851 CATCCTACGA CCGGTCCCTC AGGGTCTGGA GTATGGACAA CATGATCTGC
1901 ACGCAGACCC TGCTGCGTCA CCAGGGCAGT GTCACCGCGC TGGCTGTGTC
1951 CCGGGGCCGA CTCTTCTCAG GGGCTGTGGA TAGCACTGTG AAGGTTTGGA
2001 CTTGCTAACA GGATCCAGGC CAGGCTGTGG TTTCCCTGTA ACCAGCCCTG
2051 GACCTTTCTG AGCCAGGCTG GCCACATGGG GTGGTCTCGG GGTTTCTGCC
2101 TGCCCGTGG GCATAGGTGG ACAGGCTCTG GCAGCCGGGC AGTGCCCTCC
2151 CCGTCCCATC CTCGGCGAGC CTCCTCTAC TCGGCACTGT CTTGCTGCC
2201 CAGCCCTCTC CTGGGTGCCA GGTACGACGC TTGCCCGGCG CCACCCTCCA
2251 TCCCCACCTC CCATCCCCAC CTTAGATGGA GCGAGGGCCT TTTACTCAC
2301 CTTTCTTACC GTTTTATAGC TGTATGTAGA TTTGGTTACC TCCTGGTTGA

```

```

2351 AATAAATGCT CCACAGACTG TGGCTGTGAG TGGGGACAGC TCCTCGGGAC
2401 AAGGGGGCTG TGTGTGGCCT TGAGGTTGGT GTGCACAGGC ACTGGCTGCT
2451 GTGAGTGGGG GGGCATGGGG CAGTTTCCTT TGGTGGACCC CAGGACTTCG
2501 GCCCACTCCG GGGCCTCCCC TCCCTGCTAG GAGGCAACTC GTCACACCCA
2551 AGCTGCTGGC CTCCAGTCCC ATCTCCCCCA ACACATGTGC CCCCCAAAAG
2601 TGAGCCAGGC ACCTCTGTTT CCTGCTGTTT ATTGACAGCC GACGGCAGCG
2651 CCTTGCCAG ACCTCCCTG CCCACCTGCT GGAGCCAGC CTGTGCCGCC
2701 CTCTGAGGAG AGGCCTGGGG GGACAGCTGG GCACGTCCAC TCGCAGGGAA
2751 ACACGGGGTG AGACAGCAGG AAGGGGGCCT GCACGCCGGG ACGCCACCTC
2801 CGCCAGCCGC CTCCACCCGC CCCACACCAC AATCGCTGGT TTTCGGCATT
2851 TTTTAAATTT TTTTTTAAG AAACGTCAA GTTGTGCCCA AACTGTGGA
2901 TCAGCAAACA CGATAGAGGA GACCAGTCAG TACTTCTTGG AGGGGGCAGG
2951 AGGAGAGAGG AAAAGGGAGG GCGAGATGA CCACACAACA CAGCCTTGA
3001 CCATGAGCAG AAGCGTCCGT GGGAACTCCA CTGGGGTGA TGGGCTGCCT
3051 GCACAGCCCC TGGAGAGGGG GCCAGGCACA CCCTCAGAGG AGCTGCAAGC
3101 CCGTGGCCTG GCCTGCTACA TGCCCTGCTT CCACGTGGCT GCCACGCTGA
3151 CACACCCACA TTCACCAAAC CCACCCGCGC CCTGGGACGC AGCCACGCCA
3201 GGAGGAGGAC ACGGCCGCCG AGAGCAAGGC ACAACCTCGA GTTCTTGGGG
3251 CGCAGAGAAC TTAGGAGAGA AGCACGGAGG AGCCCCGGGC AGAGCACCCG
3301 CCCCCGGGCC CAGCCTTCC ACCTGTGCTA GCAGCCTGGG GCCTCCACTC
3351 TGGCCGGAGG AAGGACCGCA GGCAGACAGC CTGGGCTCT AACAGCTTTT
3401 GTCCGGAGCT AGACTTCGTG TCCTTTCAGT TGGTAAATGG TTTTCTATAG
3451 AATCAATAAT ATTTCTTTCT TTAATAATAT ATTTGTTAA GTTATACCTT
3501 TTTGTTTCTC TGGGGAAATC CGCCTCAGCT CATTCCTCAAT AATTAATAC
3551 TCTTGATAAA AAAAAAAAAA AGAAAAAAAA AAAA

```

BLAST Results

Entry HSBE from database EMBL:

Homo sapiens (clone exon trap d5) chromosome 16p13.3 gene, exon.
Score = 2375, P = 7.1e-101, identities = 475/475

Entry HSBD from database EMBL:

Homo sapiens (clone exon trap d32) chromosome 16p13.3 gene, exon.
Score = 876, P = 3.0e-31, identities = 176/177

Medline entries

95122486:

Structural analysis of myosin heavy chain kinase A from Dictyostelium. Evidence for a highly divergent protein kinase domain, an amino-terminal coiled-coil domain, and a domain homologous to the beta-subunit of heterotrimeric G proteins.

96149460:

Dictyostelium myosin heavy chain kinase A regulates myosin localization during growth and development.

97277316:

Identification of a protein kinase from Dictyostelium with homology to the novel catalytic domain of myosin heavy chain kinase A.

96009891:

A gene responsible for vegetative incompatibility in the fungus Podospora anserina encodes a protein with a GTP-binding motif and G beta homologous domain.

Peptide information for frame 2

ORF from 224 bp to 2005 bp; peptide length: 594
Category: similarity to known protein
Prosite motifs: ZINC_FINGER_C3HC4 (70-80)
LEUCINE_ZIPPER (436-458)
LEUCINE_ZIPPER (436-458)
G_BETA_REPEATS (335-355)
G_BETA_REPEATS (376-391)

```

1  MPPISTPRRS  DSAISVRSLSH  SESSMSLRST  FSLPEEEEEEP  EPLVFAEQPS
51  VKLCCQLCCS  VFKDPVITTC  GHTFCRRCAL  KSEKCPVDNV  KLTVVVNNIA
101  VAEQIGELFI  HCRHGCRVAG  SGKPPIFEVD  PRGCPFTIKL  SARKDHEGSC
151  DYRPVRCFNN  PSCPPLLRMN  LEAHLKECEH  IKCPHSKYGC  TFIGNQDTYE
201  THLETCTRFEG  LKEFLQQTDD  RFHEMHVALA  QKDQETAFIR  SMLGKLSEKI
251  DQLEKSLELK  FDVLDENQSK  LSEDLMEFRR  DASMLNDELS  HINARLNMGI
301  LGSYDPQQIF  KCKGTFVGHQ  GPVWCLCVYS  MGDLLFSGSS  DKTIKVWDTG
351  TTYKCQKTLE  GHDGIVLALC  IQGCKLYSGS  ADCTIIVWDI  QNLQKVNTIR
401  AHDNPNVCTLV  SSHNVLFSGS  LKAIKVWDIV  GTLKLKKELE  TGLNHWVRAL
451  VAAQSYLYSG  SYQTIKIWDI  RTLDLHVQL  TSGGSVYSIA  VTNNHIVCGT
501  YENLIHVWDI  ESKEQVRLT  GHVGTVYALA  VISTPDQTKV  FSASYDRSLR
551  VWSMDNMICT  QTLRHQGSV  TALAVSRGR  FSGAVDSTVK  VWTC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_li2, frame 2

SWISSPROT:KMH_B_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B)., N = 1, Score = 419, P = 3.6e-37

SWISSPROT:HET1_PODAN VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1., N = 1, Score = 392, P = 3.1e-33

SWISSPROT:YDJ5_SCHPO HYPOTHETICAL 67.1 KD TRP-ASP REPEATS CONTAINING PROTEIN C57A10.05C IN CHROMOSOME I., N = 1, Score = 357, P = 4.1e-30

TREMBL:AF032878_1 gene: "slimb"; product: "Slimb"; Drosophila melanogaster Slimb (slimb) mRNA, complete cds., N = 1, Score = 347, P = 1.7e-29

>SWISSPROT:KMH_B_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B).
Length = 732

HSPs:

Score = 419 (62.9 bits), Expect = 3.6e-37, P = 3.6e-37
Identities = 96/268 (35%), Positives = 158/268 (58%)

Query: 325 CLCVYSMGDLLFSGSSDKTIKVWD-TCTTYKCQKTLEGHGDIIVLALCIQGCKLYSGSADC 383
C+C +LLF+G SD +I+V+D +C +TL+GH+G V ++C L+SGS+D
Sbjct: 467 CIC----DNLLFTGCSDNSIRVYDYKSQNMCEVQTLKGHEGPVESICYNDQYLFSGSSDH 522

Query: 384 TIIVWDIQLQKVNTIRAHNPNVCTLVSSHNVLFSGSL-KAIKVWDIVGTLEKLKKELTG 442
+I VWD++ L+ + T+ HD PV T++ + LFSGS K IKVWD+ L+ K L
Sbjct: 523 SIKVWDLKKLRICFTLEGHDKPVHTVLLNDKYLFGSSDDKTIKVWDL--KTLECKYTLES 580

Query: 443 LNHWVRALVAAQSYLYSGSY-QTIKIWDI RTLDLHVQLTSGGSVYSIAVTNNHIVCGTY 501
V+ L + YL+SGS +TIK+WD++T C + L+ V +I + ++ G+Y
Sbjct: 581 HARAVKTLICISGQYLFSGSNDKTIKVWDLKTFRCNYTLKGHTKWVTTICILGTNLYSGSY 640

Query: 502 ENLIHVWDIESKEQVRLTGHVGTVYALAVISTPDQTKVFSASYDRSLRVWSMDNMICTQ 561
+ I VW+++S E TL GH V + + D+ +F+AS D +++W ++ + C
Sbjct: 641 DKTIRVWNLKSLECSATLRGHRWVEHVMVIC---DKL-LFTASDDNTIKIWDLETLCRNT 696

Query: 562 TLLRHQGSVTALAVSRGR--LFSGAVDSTVKVW 592
TL H +V LAV + + S + D +++VW
Sbjct: 697 TLEGHNATVQCLAVWEDKKCVISCSHDQSIKRVW 729

Score = 415 (62.3 bits), Expect = 1.2e-36, P = 1.2e-36
Identities = 113/303 (37%), Positives = 166/303 (54%)

Query: 255 KSLEL-KFDVLDENQSKLSEDLMEFRRDASMLNDEL-SHINARLNMGILGS-----YD 305
KS++L K ++L N+ K S +L + ++ + SH+ N+ G YD
Sbjct: 427 KSIDLEKPEILINNKKKESINLETIKLIETIKGYHVTSHLCICDNLLFTGCSDNSIRVYD 486

Query: 306 -PQQIFKCKGTFVGHQGPVWCLCVYSMGDLLFSGSSDKTIKVWDTCTTYKCQKTLEGHG 364
Q +C T GH+GPV +C Y+ LFSGSSD +IKVWD +C TLEGHG
Sbjct: 487 YKSQNMCEVQTLKGHEGPVESIC-YN-DQYLFSGSSDHSIKVWDL-KKLRICFTLEGHDK 543

Query: 365 IVLALCIQGCKLYSGSADCTIIVWDIQLQKVNTIRAHNPNVCTLVSSHNVLFSGSL-KA 423
V + + L+SGS+D TI VWD++ L+ T+ +H V TL S LFSGS K
Sbjct: 544 PVHTVLLNDKYLFGSSDDKTIKVWDLKTLECKYTLESHARAVKTLICISGQYLFSGSNDKT 603

Query: 424 IKVWDIVGTLEKLKKELTGLNHWVRALVAAQSYLYSGSY-QTIKIWDI RTLDLHVQLTSG 482
IKVWD+ + L G WV + + LYSGSY +TI++W++++L+C L+
Sbjct: 604 IKVWDL--KTFRCNYTLKGHTKWVTTICILGTNLYSGSYDKTIRVWNLKSLECSATLRGH 661

Query: 483 GGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRLTGHVGTVYALAVISTPDQTKVFS 542
 V + + + + + +N I +WD+E+ TL GH TV LAV D+ V S
 Sbjct: 662 DRWVEHVMVICDKLLFTASDDNTIKIWDLETLCRNTTLEGHNATVQCLAVWE--DKKCVIS 719

Query: 543 ASYDRSLRVW 552

S+D+S+RVW
 Sbjct: 720 CSHDQSIRVW 729

Score = 262 (39.3 bits), Expect = 3.2e-19, P = 3.2e-19
 Identities = 60/184 (32%), Positives = 109/184 (59%)

Query: 352 TYKCQKTLEGHGDIVLALCIQCGKLYSGSADCTIIVWDI--QNLQKVNTIRAHDPVCTL 409
 T K +T++G+ + LCI L++G +D +I V+D QN++ V T++ H+ PV ++
 Sbjct: 450 TIKLIETIKGYH-VTSHLCICDNLFTGCSDNSIRVYDYKSQNMCECVQLKGHEGPVESI 508

Query: 410 VSSHNVLFSGLK-AIKVWDIVGTELKKELTGLNHWVRALVAAQSYLYSGSY-QTIKI 467
 + LFSGS +IKVWD+ +L+ L G + V ++ YL+SGS +TIK+
 Sbjct: 509 CYNDQYLFSGSSDHSIKVWDL--KKLRCIFTELEGHDKPVHTVLLNDKYLFSGSSDKTIKV 566

Query: 468 WDIRTLDIHLVLTSGGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRLTGHVGTVY 527
 WD++TL+C + L++ +V ++ ++ ++ G+ + I VWD+++ TL GH V
 Sbjct: 567 WDLKTLECKYTLESHARAVKTLCSGQYLFSGSNDKTIKVWDLKTFRCNYTLKGHTKWVT 626

Query: 528 ALAVIST 534

+ ++ T
 Sbjct: 627 TICILGT 633

Score = 173 (26.0 bits), Expect = 1.7e-09, P = 1.7e-09
 Identities = 43/118 (36%), Positives = 65/118 (55%)

Query: 310 FKCKGTFVGHQGPVWCLCVYSMDLLFSGSSDKTIKVWDTCTTYKCQKTLEGHGDIVLAL 369
 F+C T GH V +C+ +G L+SGS DKTIVW+ + +C TL GHD V +
 Sbjct: 612 FRCNYTLKGHTKWVTICII--LGTNLYSGSYDKTIRVWNL-KSLECSATLRGHDRWVEHM 668

Query: 370 CIQCGKLYSGSADCTIIVWDIQLQKVNTIRAHDPV-CTLVSSHN--VLFSGLKAIKV 426
 I L++ S D TI +WD++ L+ T+ H+ V C V V+ ++I+V
 Sbjct: 669 VICDKLLFTASDDNTIKIWDLETLCRNTTLEGHNATVQCLAVWEDKKCVISCSHDQSIRV 728

Query: 427 W 427

W
 Sbjct: 729 W 729

Pedant information for DKFZphutel_1i2, frame 2

Report for DKFZphutel_1i2.2

[LENGTH] 594
 [MW] 66541.94
 [pI] 6.64
 [HOMOL] SWISSPROT:KMH_B_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B). 3e-37

[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YCR072c beta-transducin family] 2e-15
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 1e-13
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 1e-13
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 2e-11
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 2e-11
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 3e-11
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 8e-09
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YCR057c] 2e-07
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 2e-07
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 5e-07
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 5e-07

[FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 3e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YKL021c] 2e-04
 [FUNCAT] 01.03.07 deoxyribonucleotide metabolism [S. cerevisiae, YOR269w] 2e-04
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YOR212w] 0.001
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YOR212w] 0.001
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YOR212w] 0.001
 [BLOCKS] BL00678
 [BLOCKS] BL00518 Zinc finger, C3HC4 type, proteins
 [SCOP] dltbgsd 2.46.3.1.1 betal-subunit of the signal-transducing 3e-10
 [EC] 2.7.1.129 Myosin-heavy-chain kinase 3e-26
 [PIRKW] phosphotransferase 3e-26
 [PIRKW] nucleus 1e-06
 [PIRKW] plasma 9e-08
 [PIRKW] duplication 3e-25
 [PIRKW] hormone 9e-08
 [PIRKW] zinc 3e-09
 [PIRKW] cell cycle control 4e-13
 [PIRKW] transmembrane protein 3e-12
 [PIRKW] zinc finger 1e-08
 [PIRKW] stomach 9e-08
 [PIRKW] DNA binding 9e-06
 [PIRKW] autophosphorylation 3e-26
 [PIRKW] phosphoprotein 3e-26
 [PIRKW] signal transduction 5e-08
 [PIRKW] heterotrimer 5e-08
 [PIRKW] coiled coil 3e-26
 [PIRKW] multimer 3e-26
 [PIRKW] transcription regulation 4e-10
 [PIRKW] GTP binding 5e-08
 [SUPFAM] chromobox homology 9e-06
 [SUPFAM] RING finger homology 3e-09
 [SUPFAM] coatamer complex beta' chain 1e-07
 [SUPFAM] WD repeat homology 3e-26
 [SUPFAM] yeast coatamer complex alpha chain 3e-12
 [SUPFAM] GTP-binding regulatory protein beta chain 5e-08
 [SUPFAM] PRL1 protein 2e-09
 [PROSITE] WD_REPEATS 2
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MYRISTYL 14
 [PROSITE] CK2_PHOSPHO_SITE 4
 [PROSITE] ZINC_FINGER_C3HC4 1
 [PROSITE] PKC_PHOSPHO_SITE 18
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Zinc finger, C3HC4 type (RING finger)
 [PFAM] WD domain, G-beta repeats
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 6.23 %
 [KW] COILED_COIL 6.73 %

SEQ MPPISTPRRSDSAISVRSLSHSESSMSLRSTFSLPEEEEEPEPLVFAEQPSVKLCCQLCCS
 SEGXXXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXX.....
 COILS
 1gg2B

SEQ VFKDVPVITTCGHTFCRRALKSEKCPVDNVKLTVVVNNIAVAEQIGELFIHCRHGCRVAG
 SEG
 COILS
 1gg2B

SEQ SGKPPIFEVDPRGCPFTIKLSARKDHEGSCDYRPVRCPPNNPSCPPLLRMNLEAHLKECEH
 SEG
 COILS
 1gg2B

SEQ IKCPHISKYGCTFIGNQDTYETHLETCTREFGLKEFLQOTDDRFHEMHVALAQKDQEI AFLR
 SEG
 COILSCCCCCCCCCCCCCCCC
 1gg2B

SEQ SMLGKLSEKIDQLEKSLELKFVDLDENQSKLSEDLMEFRRDASMLNDELSHINARLNMG
 SEG
 COILS CCCCCCCCCCCCCCCCCCCCCCCCCC.....
 1gg2B

SEQ LGSYDPOQIFKCKGTFTVGHQGPVWCLCVYSMDLLFSGSSDKTIKVWDTCTTYKCQKLTLE
 SEG
 COILS
 1gg2BEECCCCCEEEEEETTTTCEEEEEETTTTEEEEEEG-GGCEEEEEEE

```

SEQ      GHDGIVLALCIQGCKLYSGSADCTIIVWDIQNLQKVNTIRAHDNPVCTLVSSHNVLFSGS
SEG      .....
COILS    .....
1gg2B    CCCCCEEEEETTCEEEEETTTCEEEEEETTTTEEEEE-CTTTTCCCEE.....

SEQ      LKAIKVWDIVGTGLKLKELTGLNHWVRALVAAQSYLYSGSYQTIKIWDIRTLDCIHVLQ
SEG      .....XXXXXXXXXXXXX.....
COILS    .....
1gg2B    .....

SEQ      TSGGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRTLTGHVGTVYALAVISTPDQTKV
SEG      .....
COILS    .....
1gg2B    .....

SEQ      FSASYDRSLRVWSMDNMICTQTLLRHQGSVTALAVSRGRLFSGAVDSTVKVWTC
SEG      .....
COILS    .....
1gg2B    .....

```

Prosites for DKFZphute1_i12.2

PS00001	267->271	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	50->53	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	141->144	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	340->343	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	464->467	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	330->334	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	194->200	MYRISTYL	PDOC00008
PS00008	299->305	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	364->370	MYRISTYL	PDOC00008
PS00008	379->385	MYRISTYL	PDOC00008
PS00008	419->425	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	484->490	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	568->574	MYRISTYL	PDOC00008
PS00008	583->589	MYRISTYL	PDOC00008
PS00518	70->80	ZINC_FINGER_C3HC4	PDOC00449
PS00029	436->458	LEUCINE_ZIPPER	PDOC00029
PS00678	335->350	WD_REPEATS	PDOC00574
PS00678	376->391	WD_REPEATS	PDOC00574

Pfam for DKFZphute1_i12.2

HMM_NAME WD domain, G-beta repeats

HMM *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*

++GH ++VWC+ + G + ++SGS D+T+++WD

Query 316 FVGHQGPVWCLCVYSMGDL-LFSGSSDKTIKVWD 348

22.93 519 553 1 34 dkfzphute1_i12.2 similarity to Dictostelium myosin heavy chain kinase

Alignment to HMM consensus:

Query *MrGHnnWVWCVaF..SPDGrWFivSGSWDgTCRLWD*
 ++GH ++v+++A+ +PD ++S+S D+++R+W+
 dkfzphute1 519 LTGHVGTVYALAVISTPDQTK-VFSASYDRSLRVWS 553

HMM_NAME Zinc finger, C3HC4 type (RING finger)

HMM *CPICFcTFQlDyPWPfdePmMlPCgHsFCypCIrrW..CPmC*
 C++C + F++P++++CGH+FC+ C +++ CP+
 Query 55 CQLC-----CSV---FKDPVITTCGHTFCRRCALKSEKCPVD 88

DKFZphute1_20b19

group: metabolism

DKFZphute1_20b19 encodes a novel 486 amino acid protein with similarity to bacterial sarcosine oxidases (EC 1.5.3.1.)

The novel protein seems to be a novel enzyme with sarcosine oxidase activity.

The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

similarity to sarcosine oxidases

membrane regions: 1

Summary DKFZphute1_20b19 encodes a novel 486 amino acid protein, with similarity to sarcosine oxidases.

similarity to sarcosine oxidases

complete cDNA?, complete cds potential start at Bp 48, EST hits,

Sequenced by AGOWA

Locus: unknown

Insert length: 1967 bp

Poly A stretch at pos. 1950, no polyadenylation signal found

```
1 AGCGAGGCAG CAGTGCAGCT TTCAGAGGGT CCGGGCTCAG AGGGGTTATG
51 ATTCGGAGGG TTCTGCCGCA CGGCATGGGC CGGGGCTCT TGACCCGAG
101 GCCAGGCACG CGCAGAGGAG GCTTTCTCTT GGACTGGGAT GGAAAGGTGT
151 CTGAGATTAA GAAGAAGATC AAGTCGATCC TGCCTGGAAG GTCCTGTGAT
201 CTA CTGCAAG ACACCAGCCA CCTGCCTCCC GAGCACTCGG ATGTGGTGAT
251 CGTGGGAGGT GGGGTGCTTG GCTTGTCTGT GGCCTATTGG CTGAAGAAGC
301 TGGAGAGCAG ACGAGGTGCT ATTCGAGTGC TAGTGGTGGG ACGGGACCAC
351 ACGTATTCCAC AGGCCTCCAC TGGGCTCTCA GTAGGTGGGA TTTGTCAGCA
401 GTTCTCATTT CCTGAGAACA TCCAGCTCTC CCTCTTTTCA GCCAGCTTTC
451 TACGGAACAT CAATGAGTAC CTGGCCGTAG TCGATGCTCC TCCCTGGAC
501 CTCCGGTTCA ACCCTCGGG CTACCTCTTG CTGGCTTCAG AAAAGGATGC
551 TGCAGCCATG GAGAGCAACG TGAAAGTGCA GAGGCAGGAG GGAGCCAAAG
601 TTTCTCTGAT GTCTCCTGAT CAGCTTCGGA ACAAGTTTCC CTGGATAAAC
651 ACAGAGGGAG TGGCTTTGGC GTCTTATGGG ATGGAGGACG AAGGTTGGTT
701 TGACCCCTGG TGTCTGTCTC AGGGGCTTCG GCGAAAGGTC CAGTCCCTGG
751 GAGTCCTTTT CTGCCAGGGA GAGGTGACAC GTTTTGTCTC TTCATCTCAA
801 CGCATGTTGA CCACAGATGA CAAAGCGGTG GTCTTGAAAA GGATCCATGA
851 AGTCCATGTG AAGATGGACC GCAGCCTGGA GTACCAGCCT GTGGAATGCG
901 CCATTGTGAT CAACGCAGCC GGAGCCTGGT CTGCGCAAAT CGCAGCACTG
951 GCTGGTGTG GAGAGGGGCC GCCTGGCACC CTGCAGGGCA CCAAGCTACC
1001 TGTGGAGCCG AGGAAAAGGT ATGTGTATGT GTGGCACTGC CCCAGGGAC
1051 CAGGCCTAGA GACTCCGCTT GTTGACAGCA CCAGTGGAGC CTATTTTCGC
1101 CGGAAGGAT TAGGTAGCAA CTACCTAGGT GGTCTAGACC CCACTGAGCA
1151 GGAAGAACC GACCCGGCGA ACCTGGAAGT GGACCATGAT TTCTTCCAGG
1201 ACAAGGTGTA GCCCCATTG GCCCTGAGGG TCCCAGCTTT TGAGACTCTG
1251 AAGGTTGAGA GCGCTGGGC CGGCTATTAC GACTACAACA CCTTTGACCA
1301 GAATGGCGTG GTGGGCCCCC ACCCGCTAGT TGTCAACATG TACTTTGCTA
1351 CTGGCTTCAG TGGTCACGGG CTCCAGCAGG CCCCTGGCAT TGGGCGAGCT
1401 GTAGCAGAGA TGGTACTGAA GGGCAGGTTT CAGACCATCG ACCTGAGCCC
1451 CTTCCTCTTT ACCCGCTTTT ACTTGGGAGA GAAGATCCAG GAGAACAAACA
1501 TCATCTGAGC ATGTGTGCTC TGCACTGGCT CCACTGGCTT GCATCCTGGC
1551 TGTGTTACCA GCCTGTGTTG CTGCTTCCAT CTTCCCAGT ACTGTGCCAG
1601 GCCTTCTCCC CCTCCCAGT GTCTCTCCT CTCAGGCAGG CCATTGCACC
1651 CATATGGCTG GGCAGGCACA GGCAGTGAGG CCGAGGCCAA TAGCGAGTGA
1701 TGAGCGGGAT CCTAGGACTG ATCTGTAGCC CATGTGATG TCACCCACCA
1751 GGGCAATCCA TCTGGAGGCC TGAGCACCTT GGCCAGGAC TGGCTTCATC
1801 CTGGCACTGA CCAGGAAAGA CTGCCTCTGA CCCTCTTAGC AGACAGAGCC
1851 CAGGCATGGG AGCACTCTGG GGCAGCCTGG CTCAGGTTTA TTGATTTTCG
1901 TCTGTTTACC CTATCCATTA ATCAATACAT GTAATTAAC CCTTCCCTCC
1951 AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 48 bp to 1505 bp; peptide length: 486
 Category: similarity to known protein

```

1 MIRRVLPHGM GRGLLTRPG TRRGGSFDW DGKVSEIKKK IKSILPGRSC
51 DLLQDTSHLP PEHSDVVIVG GVLGLSVAY WLKKLESRRG AIRVLVVERD
101 HTYSQASTGL SVGGICQQFS LPENIQLSLF SASFLRNINE YLAVVDAPPL
151 DLRFNPSGYL LLASEKDAAA MESNVKVRQ EGAKVSLMSP DQLRNKFPWI
201 NTEGVALASY GMEDEGWFDW WCLLQGLRRK VQSLGVLFCQ GEVTRFVSSS
251 QRMLTTDDKA VVLKRIHEVH VKMDRSLEYQ PVECAIVINA AGAWSAQIAA
301 LAGVGEGPPG TLQGTCLPVE PRKRYVYVWH CPQGPGLTLP LVADTSGAYF
351 RREGLGSNYL GGRSPTEQEE PDPANLEVDH DFFQDKVWPH LALRVPAFET
401 LKVQSAWAGY YDNTFDQNG VVGPHPLVVN MYFATGFSGH GLQAPGIGR
451 AVAEMVLKGR FQTIDLSFPL FTRFYLGSKI QENNII

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phut1_20b19, frame 3

TREMBL:CEM04B2_4 gene: "M04B2.4"; *Caenorhabditis elegans* cosmid M04B2,
 N = 1, Score = 801, P = 9.2e-80

PIR:B71184 probable sarcosine oxidase - *Pyrococcus horikoshii*, N = 2,
 Score = 194, P = 2e-26

PIR:B69284 sarcosine oxidase, subunit beta (soxB) homolog -
Archaeoglobus fulgidus, N = 3, Score = 189, P = 8.2e-22

TREMBL:AF042732_1 gene: "Bb"; product: "unknown protein"; *Anopheles gambiae* (Bb) gene, partial cds; and TU37B2 (TU37B2) and diphenol oxidase-A2 (Dox-A2) genes, complete cds., N = 1, Score = 386, P = 8.7e-36

PIR:F71008 probable sarcosine oxidase - *Pyrococcus horikoshii*, N = 2,
 Score = 200, P = 4e-25

>TREMBL:CEM04B2_4 gene: "M04B2.4"; *Caenorhabditis elegans* cosmid M04B2
 Length = 527

HSPs:

Score = 801 (120.2 bits), Expect = 9.2e-80, P = 9.2e-80
 Identities = 171/433 (39%), Positives = 260/433 (60%)

```

Query:      61 PEHSDVVIVGGVGLSVAYWLKKLESRRGAIRVLVVERDHTYSQASTGLSVGGICQQFS 120
           P  +++VI+GGG+ G S A+WLK+  R  +V+VVE +  ++++ST LS GGI QQFS
Sbjct:      91 PYRAEIVIIIGGGLSGSSTAFWLKE-RFRDEDFKVVVVENNDVFTKSSTMLSTGGITQQFS 149

Query:     121 LPENIQLSLFSASFLRNINEYLAVVDAPPLDLRFNPSGYLLA-SEKDAAAMESNVKVR 179
           +PE + +SLF+ FLR+  E+L ++D+  D+ F P+GYL LA ++++  M S KVQ
Sbjct:     150 IPEFVDMSLFTTEFLRHAGEHLRIIDSEQPDINFPTGYLRLAKTDEEVEMMRSAWKVQI 209

Query:     180 QEGAKVSLMSPDQLRNKFPWINTGVALASYGMEDEGWFDWCLLQGLRRKVQSLGVLF 239
           + GAKV L+S D+L  ++P++N + V LAS G+E+EG  D W LL  +R K  +LGV +
Sbjct:     210 ERGAKVQLLSKDELTKRYPYMNVDVLLASLGVENEGTIDTWQLLSAIREKNITLGVQYV 269

Query:     240 QGEVTRFVSSSQRM-----LTDDKAVVLKRIHEVHVKMDRS-LEYQPVECAIVI 288
           +GEV F  R  T D+  + +RI V V+  +  +P+  +++
Sbjct:     270 KGEVEGFQFERHRASSEVHAFGDDATADENKLAQRISGVLVRPQMNDASARPIRAHLIV 329

Query:     289 NAAGAWSAQIAALAGVGEGPPGTTLQGTCLPVEPRKRYVYVWHCPQGPGLTLPVADTS-G 347
           NAAG W+ Q+A +AG+G+G  G L  +P++PRKR V+V  P  P  + P + D S G
Sbjct:     330 NAAGWAGQVAKMAGIGGT-GLL-AVVPVIQPRKRDVVFIFAPDVPS-DLPFIIDPSTG 386

Query:     348 AYFRREGLGSNYLGGSPTEQEEP--DPANLEVDHDFQDKVWPHLALRVPAFETLKVQS 405
           + R+  G  +L GR+P+++E+  D +NL+VD+D F  K+WP L  RVP F+T KV+S
Sbjct:     387 VECRQTDGQTFLVGRTPSKEEDAKRDHSLNDVDYDDFYQKIWFPVLVDRVPFGFQAKVKS 446

```

Pedant information for DKFZphutel_20b19, frame 3

Report for DKFZphute1 20b19.3

```

SEQ      MIRRVLPHGMRGLLTRRPGRTRGGFSLDWDGKVS EIKKKIKSILPGRSCDLLQDTS HLP
SEG      .....XXXXXXXXXXXXXXXXX.....XXXXXXXXX
PRD      cccceccccccceecccccccccccccccccchhhhhhhhhccccccccceeccccccc
MEM      .....

SEQ      PEHSDVIVVGGGVLGLSVAYWLKKLESRRGAIRVLVVERDHTYSQASTGLSVGGICQQFS
SEG      .....XXXXXXXXXXXXX.....
PRD      cccceeeccccchhhhhhhhhhhhhhhccccceeeeeeccccccccccccccccccccceec
MEM      .....MMMMMMMMMMMMMMMMM.....

SEQ      LPENIQLSLFSASFLRNINEYLAVVDAPPLDLRFNPSGYLLASEKDAAAMESNVKVQRQ
SEG      .....
PRD      cchhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeeccccceeehhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      EGAKVSLMSPDQLRNKFPWINTEGVALASYGMEDEGWFDPPWCLLQGLRRKVQSLGVLFCQ
SEG      .....
PRD      cccceeeccccchhhhhhhccccccccccccccccccccccccchhhhhhhhhhhheeeec
MEM      .....

SEQ      GEVTRFVSSQRMLTTDDKAVVLKRIHEVHVKMDRSLEYQPVECAIVINAAGAWSAQIAA
SEG      .....
PRD      ceeeeccccccccccccchhhhhhhhhheeeccccccccceeeeeeccccchhhhhh
MEM      .....

SEQ      LAGVGEGPGPTLQGTKLPVEPRKRYVYVWHCPQGPGLETPLVADTSGAYFRREGLGSNYL
SEG      .....
PRD      hhccccccccccccccccccccceeeeeeccccccccceeeccccceeeccccccce
MEM      .....

SEQ      GGRSPTEQEEDPANLEVDHDFQDKVWPHLALRVAPAFETLKVQSAWAGYYDYNTFDQNG
SEG      .....
PRD      eccccccccccccccccccccchhhhhhhhhhhhhhhcchhhhhhhhhhhheeeccccccc
MEM      .....

SEQ      VVGPHPLVVNMVFATGFSGHGLQQA PGIGRAVAVMLKGRFQITDLSPLFLTRFYLGEXI
SEG      .....
PRD      cccccccccceeeccccccccccccchhhhhhhhhhhhhhhccccceeecccccccccccccc
MEM      .....

SEQ      QENNII
SEG      .....
PRD      ccccccc
MEM      .....

```

Prosites for DKFZphut1_20b19.3

PS00002	438->442	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	87->90	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	400->403	PKC_PHOSPHO_SITE	PDOC00005
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	364->368	CK2_PHOSPHO_SITE	PDOC00006
PS00006	366->370	CK2_PHOSPHO_SITE	PDOC00006
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	109->115	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00008	354->360	MYRISTYL	PDOC00008
PS00008	447->453	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_20b19.3)

DKFZphut1_20g21

group: signal transduction

DKFZphut1_20g21 encodes a novel 861 amino acid protein with partial similarity to human ras inhibitor and other ras inhibitor proteins.

Ras is a signal transducing molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. The novel protein seems to be a new ras inhibitor protein.

The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

Ras inhibitor

additional 1188 Bp at 5' and 1107 at 3' end in comparison to I22483

Sequenced by AGOWA

Locus: unknown

Insert length: 4137 bp

Poly A stretch at pos. 4116, no polyadenylation signal found

```

1 GGGAGAAGT AAACAGGAGA TGGTGGCGAC AGATGTCAAC CTGGAAAATG
51 GCCTGGAACC CGCTGAAACC CACAGCATGG TAAGACACAA GGATGGTGGC
101 TATTCCGAGG AAGAGGACGT GAAGACCTGT GCCCGGGACT CAGGCTATGA
151 CAGCCTCTCC AACAGGCTCA GCATCTTGA CCGGCTCCTC CACACCCACC
201 CCATATGGCT GCAGCTGAGT CTGAGTGAGG AGGAGGCAGC AGAGGTCCTG
251 CAGGCCCCAGC CTCGGGGGAT CTCTCTGGTT CATAAATCTA CCAAGATGCA
301 GAAGAAAGTC CTCTCCCTCC GCCTGCCCTG TGAATTGGG GCCCCTCA
351 AGGAATTTGC CATAAAGGAA AGCACATACA CCTTTTCCCT GGAAGGCTCA
401 GGAATCAGTT TCGCAGATTT ATTCCGGCTC ATGTCTTCT ACTGCATCAG
451 CAGGGATGTT CTACCATTTA CCTTGAAGTT GCCTTATGCC ATTTCAACAG
501 CCAAGTCGGA GGCTCAGCTT GAAGAACTGG CCCAGATGGG ACTAAATTTT
551 TGGAGCTCCC CAGCTGACAG CAAACCCCGG AACCTTCCAC CTCCCCATAG
601 GCCTCTTTCC TCCGACGGTG TCTGTCTGCT CTCCTGCGT CAGCTCTGCC
651 TTATAAATGG AGTGCAATTCT ATCAAAACCA GGACGCCTTC AGAGCTGGAG
701 TGCAGCCAGA CCAACGGGGC CCTGTGCTTT ATTAATCCCC TTTTCTTGAA
751 AGTGACACAG CAGGACCTCA GTGGAGGCCT GAAACGGCCG AGCACAAGGA
801 CTCCCAACGC GAATGGCAGC GAGCGGACTC GGTCCCCCCC ACCCAGGCCC
851 CCGCCACCCG CTATTAATAG TCTCCACACA AGCCCTCGGC TGGCCAGGAC
901 TGAAACCCAG ACGAGCATGC CAGAAACAGT CAACCATAA CACATGCGGA
951 ACGTAGCTCT GCCTGGAACG AAACCAATCT CCATCCCTCC ACCCGGCTG
1001 AAGAAGCAGG CTTCTTTTCT GGAAGCAGAG GGCGGTGCAA AGACCTTGAG
1051 CCGCGGCCGG CCGGGCGCAG GCCCGGAGCT GGAGCTGGGC ACAGCTGGCA
1101 GCCCAGGTGG GGGCCCGCCT GAGGCCGCCC CGGGGGATTG CACAAGGGCC
1151 CCGCGGCCCA GCTCTGAATC ACGGCCCCCG TGCCATGGAG GCGGGCAGCG
1201 GCTGAGCGAC ATGAGCATTT CTAATCTCTC CTCCGACTCG CTGGAGTTCT
1251 ACCGGAGCAT GCCTCTGTTT GGCTACGAGG CGGACACCAA CAGCAGCCTG
1301 GAGGACTACG AGGGGGAAG TGACCAAGAG ACCATGGCGC CCCCCATCAA
1351 GTCCAAAAG AAAAGGAGCA GCTCCTTCGT GCTGCCAAG CTGCTCAAGT
1401 CCCAGCTGCA GAAGGTGAGC GGGGTGTTCA GCTCCTTCAT GACCCCGGAG
1451 AAGCGGATGG TCCGCAGGAT CGCCGAGCTT TCCCGGGACA AATGCACCTA
1501 CTTCCGGTGC TTAGTGACAG ACTACGTGAG CTTCTGCGAG GAGAACAAAG
1551 AGTGCCACGT GTCCAGCACC GACATGCTGC AGACCATCCG GCAGTTCATG
1601 ACCCAGGTCA AGAACTATTT GTCTCAGAGC TCGGAGCTGG ACCCCCCCAT
1651 CGAGTCGCTG ATCCCTGAAG ACCAAATAGA TGTGGTGCTG GAAAAAGCCA
1701 TGCACAAGTG CATCTTGAAG CCCCTCAAGG GGCATGTGGA GGCCATGCTG
1751 AAGGACTTTC ACATGGCCGA TGGCTCATGG AAGCAACTCA AGGAGAACCT
1801 GCAGCTTGTG CGGCAGAGGA ATCCGCGAGG GCTGGGGGTC TTCGCCCCGA
1851 CCCCTGATTT TGTGGATGTG GAGAAAATCA AAGTCAAGTT CATGACCATG
1901 CAGAAGATGT ATTCGCCGGA AAAGAAGGTC ATGCTGCTGC TCGGGGTCTG
1951 CAAGTCTATT TACACGGTCA TGGAGAACAA CTCAGGGAGG ATGTATGGCG
2001 CTGATGACTT CTTGCCAGTC CTGACCTATG TCATAGCCCA GTGTGACATG
2051 CTTGAATTGG AACTGAAAT CGAGTACATG ATGGAGCTCC TAGACCCATC
2101 GCTGTTACAT GGAGAAGGAG GCTATTACTT GACAAGCGCA TATGGAGCAC
2151 TTTCTCTGAT AAAGAATTTC CAAGAAGAAC AAGCAGCGCG ACTGCTCAGC
2201 TCAGAAACCA GAGACACCTT GAGGCACTGG CACAAACGGA GAACCCACAA
2251 CCGGACCATC CCCTCTGTGG ACGACTTCCA GAATTACCTC CGAGTTGCAT
2301 TTCAGGAGGT CAACAGTGGT TGCACAGGAA AGACCTCCTT TGTGAGACCT
2351 TACATCACCA CTGAGGATGT GTGTGAGATC TCGCTGAGA AGTTCAAGGT
2401 GGGGGACCTT GAGGAGTACA GCCTCTTTCT CTTCTGTTGAC GAGACATGGC
2451 AGCAGCTGCG AGAGGACACT TACCTCAAAA AAATCAAGGC GGAGCTGCAC

```

```

2501 AGCCGACCAC AGCCCCACAT CTTCCACTTT GTCTACAAAC GCATCAAGAA
2551 CGATCCTTAT GGCATCATT TCCAGAACGG GGAAGAAGAC CTCACCACCT
2601 CCTAGAAAGAC AGGCGGGACT TCCAGTGGT GCATCCAAAG GGGAGCTGGA
2651 AGCCTTGCCCT TCCCGCTTCT ACATGCTTGA GCTTGAAAAG CAGTCACCTC
2701 CTGGGGGACC CCTCAGTGTA GTGACTAAGC CATCCACAGG CCAACTCGGC
2751 CAAGGGCAAC TTTAGCCACG CAAGGTAGCT GAGGTTTGTG AAACAGTAGG
2801 ATTCTCTTTT GGCAATGGAG AATTGCATCT GATGGTTCAA GTGTCCTGAG
2851 ATTGTTTGCT ACCTACCCCC AGTCAGGTTT TAGGTTGGCT TACAGGTATG
2901 TATATGTGCA GAAGAAACAC TTAAGATACA AGTCTTTTGT AATTCAACAG
2951 CAGATGCTTG CGATGCAGTG CGTCAGGTGA TTCTCACTCC TGTGGATGGC
3001 TTCATCCCTG CCTTCCTTCC TTTCTTTTTC CTTTTTTTTT TTTTTTTTTT
3051 TTTTACAAA GAGCCTTCAT GTTTTTATAT ATTCATAGA AATTTTATA
3101 GCAGTTGCAG GTAAACTGTC AGGATTGGTT TTAAATATT TTTGTAACCT
3151 TAAATATTTC TATAATTATG CATGTGATTT TAACATTAA TATCAAAAA
3201 TAAATCTCTT GCTGGATTG AGAGTATTGC ATTTTAAAG TCTCTCTTCT
3251 GTAACTGGAT GTTTTGGCAA CTTTGTGGG AGAGACTGCT GGATTTCTTA
3301 AAGCAACGTA TTCTGACAC TGGCCACAGA ATGCTTTGG AAATCGGATG
3351 TACTGTTCTC TTGTTACCGT TTAGTGGTGT TTTGCTGTTT TGTTTTTAA
3401 ACAAAATGAT CTGAGAATAA GGAGAGAAAT GAATGTAGAG AGAGGTAGAG
3451 AGAGAAATAT GAACTCTAAC AAAGGACTGA GGAGTGCAGT CTGCTGGTTC
3501 AGGCTCTTCA AAAGATGTAG AAAAAGAGAT AGAAGGAACC ACCTATGCTT
3551 AAAATACTGT AAATATGCAG TGAGGTTTGG CAAAATCTAT TCCATGTGTG
3601 ATTTGCTTGT AGAAACAATT TTGAAAGCCC CTTGAGGAAA ATAAAAATCA
3651 AGAAGAACAC TTTTCTCCCT TTTCCATACA AATTAAACT TAACAGCATC
3701 AAATTATTGG GACCAGAAAC CAAGTAATGT ATAATGTGGC TTTTGTGAG
3751 TTAATAAAGA TGCTATATAA TGGAGAAGAA TTTGAAAATG CACAAAAAAA
3801 TCAATCTACA TTATCAGAAC CTGCAGTGAA ATTAACTTA TGTAAATAA
3851 AACCAGTTTG CAGGTGCACA AACTATGAGG GTCTTGATC CACGTAACAC
3901 AGGTAGTTAC AAAAACATGT TATTGTACTG TGTAAAGATG CATAGTCATC
3951 TCATTTGGTT GGCTTTGTAC CTTGTACCTT TTTTAGCCTT GGCTTTGTG
4001 GAACTAGAAC CCTCAGCACA TACTGTGTTG TACTTTTGTA AATGATTTT
4051 TAAATGGAAT TTGACACATA ATACATTGTA ATACTGTATG ATAATCATGT
4101 GTGAAATAA TTTTGAAT AAAAATAA AAAAAA

```

BLAST Results

Entry I22483 from database EMBL:
Sequence 15 from patent US 5527896.
Length = 1829
Plus Strand HSPs:
Score = 9097 (1364.9 bits), Expect = 0.0, P = 0.0
Identities = 1821/1823 (99%), Positives = 1821/1823 (99%),

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2602 bp; peptide length: 861
Category: known protein
Classification: Cell signaling/communication

```

1  MVRTDVNLEN GLEPAETHSM VRHKDGGYSE EEDVKTCARD SGYDSLNRLL
51  SILDRLHHTH PIWLQLSLSE EEAAEVLQAO PPGIFLVHKS TKMQKKVLSL
101 RLPCEFGAPL KEFAIKESTY TFSLESGSIS FADLFRLIAF YCISRDVLPF
151 TLKLPYAIST AKSEAQLEEL AQMGLNFWSS PADSKPPNLP PPHRPLSSDG
201 VCPASLRQLC LINGVHSIKT RTPSELECSQ TNGALCFINP LFLKVHSQLD
251 SGGGLKRPSTR TPNANGTERT RSPPRP PPPP AINSHTSPR LARTETQISM
301 PETVNHKNGH NVALPGTKPT PIPPPRLKKQ ASFLEAEGGA KTLSSGRPGA
351 GPELELGTAG SPGGAPPEAA PGDCTRAPP SSESPPCHG GRQRLSDMSI
401 STSSDSLEF DRSMPLFGYE ADTNSLEDY EGESDQETMA PPIKSKKKRS
451 SSFVLPKLKVK SQLQKVSGVF SSFMTPEKRM VRRIAELSRD KCTYFGCLVQ
501 DYVSFLQENK ECHVSSDML QTIRQFMQTV KNYLSQSSEL DPPIESLIPE
551 DQDQDVLEKA MHKCILKPLK GHVEAMLKDF HMDGSKWQL KENLQVLRQR
601 NPQELGVFAP TPDFVDVEKI KVKFMTMQKM YSPEKKVMML LRVCKLIYTV
651 MENNSGRMYG ADDFLPVLTY VIAQCDMLEL DTEIYMMEL LDPSLLHGEG
701 GYXLTSAIGA LSLIKNFQEE QAARLLSSET RDTLRQWHKR RTNRTIPSV
751 DDFQNYLRVA FQEVNSGCTG KTLVVRPYIT TEDVCQICAE KFKVGDPEEY
801 SLFLFVDETW QQLAEDTYPQ KIKAEHLSRP QPHIFHFVYK RIKNDPYGII
851 FQNGEEDLTT S

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_20g21, frame 2

TREMBL:RNU80076_1 product: "RIN1"; Rattus norvegicus RIN1 mRNA, complete cds., N = 3, Score = 606, P = 6.8e-97

PIR:A38637 Ras interactor RIN1 - human, N = 3, Score = 587, P = 1.9e-92

TREMBL:HSRASINL_1 product: "ras inhibitor"; Human ras inhibitor mRNA, 3' end., N = 2, Score = 592, P = 9.8e-61

SWISSPROT:RIN1_HUMAN RAS INTERACTION/INTERFERENCE PROTEIN 1 (RAS INHIBITOR JC99) (FRAGMENT)., N = 2, Score = 587, P = 4.1e-60

PIR:B38637 Ras inhibitor (clone JC265) - human (fragment), N = 1, Score = 2446, P = 4.6e-254

>PIR:B38637 Ras inhibitor (clone JC265) - human (fragment)
Length = 471

HSPs:

Score = 2446 (367.0 bits), Expect = 4.6e-254, P = 4.6e-254
Identities = 471/471 (100%), Positives = 471/471 (100%)

```

Query:   391 GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS 450
          GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS
Sbjct:   1  GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS 60

Query:   451 SSFVLPKLVKSQLQKVSGVSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 510
          SSFVLPKLVKSQLQKVSGVSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK
Sbjct:   61 SSFVLPKLVKSQLQKVSGVSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 120

Query:   511 ECHVSSDMLQTIQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK 570
          ECHVSSDMLQTIQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK
Sbjct:   121 ECHVSSDMLQTIQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK 180

Query:   571 GHVEAMLKDFHMADGSKQLKENLQLVRQRNPQELGVFAPTDFVDVEKIKVKFMTMQKM 630
          GHVEAMLKDFHMADGSKQLKENLQLVRQRNPQELGVFAPTDFVDVEKIKVKFMTMQKM
Sbjct:   181 GHVEAMLKDFHMADGSKQLKENLQLVRQRNPQELGVFAPTDFVDVEKIKVKFMTMQKM 240

Query:   631 YSPEKKVMLLLRVCKLIYTMENNSGRMYGADDFLPVLTYYVIAQCDMLELDEIEYMMEL 690
          YSPEKKVMLLLRVCKLIYTMENNSGRMYGADDFLPVLTYYVIAQCDMLELDEIEYMMEL
Sbjct:   241 YSPEKKVMLLLRVCKLIYTMENNSGRMYGADDFLPVLTYYVIAQCDMLELDEIEYMMEL 300

Query:   691 LDPSLLHGEGGYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRITNRTIPSV 750
          LDPSLLHGEGGYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRITNRTIPSV
Sbjct:   301 LDPSLLHGEGGYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRITNRTIPSV 360

Query:   751 DDFQNYLRVAFQEVNSGCTGKTLVVRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW 810
          DDFQNYLRVAFQEVNSGCTGKTLVVRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW
Sbjct:   361 DDFQNYLRVAFQEVNSGCTGKTLVVRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW 420

Query:   811 QQLAEDTYPQKIKAEHLSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS 861
          QQLAEDTYPQKIKAEHLSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS
Sbjct:   421 QQLAEDTYPQKIKAEHLSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS 471

```

Pedant information for DKFZphut1_20g21, frame 2

Report for DKFZphut1_20g21.2

```

[LENGTH]      861
[MW]           96380.26
[pI]           6.15
[HOMOL]       PIR:B38637 Ras inhibitor (clone JC265) - human (fragment) 0.0
[FUNCAT]      08.13 vacuolar transport [S. cerevisiae, YML097c] 3e-10
[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]
3e-10
[FUNCAT]      30.03 organization of cytoplasm [S. cerevisiae, YML097c] 3e-10
[FUNCAT]      08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]
3e-10
[PIRKW]       alternative splicing 3e-59
[SUPFAM]      Ras interactor RIN1 3e-59

```

(No Prosite data available for DKFZphute1_20g21.2)
(No Pfam data available for DKFZphute1_20g21.2)

DKFZphutel_20h13

group: intracellular transport and trafficking

DKFZphutel_20h13 encodes a novel 955 amino acid protein with similarity to alpha-adaptins.

Adaptins are components of the adaptor complexes which link clathrin to receptors in coated vesicles. The alpha-adaptins, which are found exclusively in endocytic coated vesicles, separate into two bands on SDS gels, designated A and C. The novel protein is very similar to both alpha adaptin A and C. The novel protein is a new human alpha-adaptin.

The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

strong similarity to alpha-adaptins

complete cDNA, complete cds start at Bp 78, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3352 bp

Poly A stretch at pos. 3297, polyadenylation signal at pos. 3279

```

1 GCGCCCCGTC CCCGCTTGCC AGCCCCCGCT GCTCTGTGCC CTGTCCGGCC
51 AGGCCTGGAG CCGACACCAC CGCCATCATG CCGGCCGTGT CCAAGGGCGA
101 TGGGATCGCG GGGCTCGCGG TGTTTCATCTC CGACATCCGG AACTGTAAGA
151 GCAAAGAGGC GGAAATTAAG AGAATCAACA AGGAACTGGC CAACATCCGC
201 TCCAAGTTCA AAGGAGACAA AGCCTTGGAT GGCTACAGTA AGAAAAATA
251 TGTGTGTAAA CTGCTTTTCA TCTTCTGTCT TGGCCATGAC ATTGACTTTG
301 GGCACATGGA GGCTGTGAAT CTGTTGAGTT CCAATAAATA CACAGAGAAG
351 CAAATAGGTT ACCTGTTTCA TTCTGTGCTG GTGAACTCGA ACTCGGAGCT
401 GATCCGCGCT ATCAACAACG CCATCAAGAA TGACCTGGCC AGCCGCAACC
451 CCACCTTCAT GTGCCTGGCC CTGCACTGCA TCGCCAACGT GGGCAGCCGG
501 GAGATGGGCG AGGCCCTTTC CGCTGACATC CCCCAGCATC TGGTGGCCGG
551 GGACAGCATG GACAGTGTCA AGCAGAGTGC GGCCCTGTGC CTCCTTCGAC
601 TGTACAAGGC CTCGCTGAC CTGGTGCCCA TGGCGGAGTG GACGGCGCGT
651 GTGGTACACC TGCTCAATGA CCAGCACATG GGTGTGGTCA CGGCCGCCGT
701 CAGCCTCATC ACCTGTCTCT GCAAGAAGAA CCCAGATGAC TTCAAGACGT
751 GCGTCTCTCT GGCTGTGTCT CGCCTGAGCC GGATCGTCTC CTCTGCCTCC
801 ACCGACCTCC AGGACTACAC CTACTACTTC GTCCCAGCAC CCTGGCTCTC
851 GGTGAAGCTC CTGCGGCTGC TGCAGTGCTA CCGCCTCCA GAGGATGCCG
901 CTGTGAAGGG GCGGCTGGTG GAATGTCTGG AGACTGTGCT CAACAAGGCC
951 CAGGAGCCCC CCAATCCAA GAAGGTGCAG CATTCCAACG CCAAGAACGC
1001 CATCCTCTTC GAGACCATCA GCCTCATCAT CCACTATGAC AGTGAGCCCA
1051 ACCTCCTGGT TCGGGCCTGC AACCAGCTGG GCCAGTTTCT GCAGCACCAG
1101 GAGACCAACC TCGGCTACCT GGCCCTGGAG AGCATGTGCA CGCTGGCCAG
1151 CTCCGAGTTC TCCCATGAAG CCGTCAAGAC GCACATTGAC ACCGTCATCA
1201 ATGCCCTCAA GACGGAGCGG GACGTCAGCG TCGCGCAGCG GCGCGCTGAC
1251 CTCCTCTACG CCATGTGTGA CCGGAGCAAT GCCAAGCAGA TCGTGTCCGA
1301 GATGCTGCGG TACCTGGAGA CGGCAGACTA CGCCATCCGC GAGGAGATCG
1351 TCCTGAAGGT GGCCATCTCT GCCGAGAAGT ACGCCGTGGA CTACAGCTGG
1401 TACGTGGACA CCATCCTCAA CCTCATCCGC ATTGCGGGCG ACTACGTGAG
1451 TGAGGAGGTG TGGTACCGTG TGCTACAGAT CGTCACCAAC CGTGATGACG
1501 TCCAGGGGTA TGCCGCCAAG ACCGTCTTTG AGGCGCTCCA GGCCCTTGCC
1551 TGTACAGAGA ACATGGTGAA GGTGGCGGCG TACATCCTTG GGGAGTTTGG
1601 GAACCTGATT GCTGGGGACC CCGCTCCAG CCCCCAGTG CAGTTCCTCC
1651 TGCTCCACTC CAAGTTCCAT CTGTGCAGCG TGGCCACGCG GGCGCTGTCT
1701 CTGTCCACCT ACATCAAGTT CATCAACCTC TTCCCGAGA CCAAGGCCAC
1751 CATCCAGGGC GTCTGCGGG CCGGCTCCA GCTGCGCAAT GCTGACGTGG
1801 AGCTGCAGCA GCGAGCCGTG GAGTACCTCA CCCTCAGCTC AGTGGCCAGC
1851 ACCGACGTCC TGGCCACGGT GCTGGAGGAG ATGCCGCCCT TCCCGAGCG
1901 CGAGTCGTCC ATCCTGGCCA AGCTGAAACG CAAGAAGGGG CCAGGGGCCG
1951 GCAGCGCCCT GGACGATGGC CGGAGGGACC CCAGCAGCAA CGACATCAAC
2001 GGGGCGATGG AGCCACCCC CAGCACTGTG TCGAGCCCTC CGCCCTCCGC
2051 CGACCTCTCT GGGCTGCGGG CAGCCCTTCC CCGGCGAGCA CCCCCGGCT
2101 CTGCAAGGAC AGGGAACCTT CTGGTGGACG TCTTCGATGG CCGGCGCGCC
2151 CAGCCAGGCC TGGGGCCAC CCCCAGGAG GCCTTCTCTA GCCCAGGTCC
2201 TGAGGACATC GGCCCTCCCA TTCCGGAAGC CGATGAGTTG CTGAATAAGT
2251 TTGTGTGTAA GAACAACGGG GTCCTGTTTC AGAACCAGCT GCTGCAGATC
2301 GGAGTCAAGT CAGAGTTCCG ACAGAACCTG GGCCGCATGT ATCTCTCTTA
2351 TGGCAACAAG ACCTCGGTGC AGTTCAGAAA TTTCTCACC ACTGTGGTTC
2401 ACCCGGGAGA CCTCCAGACT CAGCTGGCTG TGCAGACCAA GCGCGTGGCG
2451 GCGCAGGTGG ACGGCGGCGC GCAGGTGCAG CAGGTGCTCA ATATCGAGTG
2501 CCTGCGGGAC TTCCTGACGC CCCCCTGCT GTCCGTGCGC TTCCGGTACG
2551 GTGGCGCCCC CCAGGCCCTC ACCCTGAAGC TCCAGTGAC CATCAACAAG

```

```

2601 TTCTTCCAGC CCACCGAGAT GCGGCCCCAG GATTCTCTCC AGCGCTGGAA
2651 GCAGCTGAGC CTCCCTCAAC AGGAGGCGCA GAAATCTTC AAAGCCAACC
2701 ACCCCATGGA CGCAGAAGTT ACTAAGGCCA AGCTTCTGGG GTTTGGCTCT
2751 GCTCTCTGGG ACAATGTGGA CCCCAACCCT GAGAACTTCG TGGGGGCGGG
2801 GATCATCCAG ACTAAAGCCC TGCAGGTGGG CTGTCTGCTT CGGCTGGAGC
2851 CCAATGCCCA GGCCCGAGATG TACCGGCTGA CCCTGCGCAC CAGCAAGGAG
2901 CCCGTCTCCC GTCACCTGTG TGAGCTGCTG GCACAGCAGT TCTGAGCCCT
2951 GGA CTCTGCC CCGGGGGATG TGGCCGGCAC TGGGCAGCCC CTTGGACTGA
3001 GGCAGTTTGT GTGGATGGGG GACCTCCACT GGTGACAGAG AAGACACCAG
3051 GGTTTGGGGG ATGCCTGGGA CTTTCCTCCG GCCTTTTGTA TTTTATTTT
3101 TGTTTCATCTG CTGCTGTTTA CATTCTGGGG GGTAGGGGG AGTCCCCCTC
3151 CCTCCCTTTC CCCCCAAGC ACAGAGGGGA GAGGGGCCAG GGAAGTGGAT
3201 GTCTCCTCCC CTCCCACCCC ACCCTGTTGT AGCCCTCCTT ACCCCTCCC
3251 CATCCAGGGG CTGTGTATTA TTGTGAGCGA ATAAACAGAG AGACGCTAAA
3301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AA

```

BLAST Results

No BLAST result

Medline entries

89155572:

Cloning of cDNAs encoding two related 100-kD coated vesicle proteins (alpha-adaptins).

97431776:

Alpha-adaptin, a marker for endocytosis, is expressed in complex patterns during *Drosophila* development.

Peptide information for frame 3

ORF from 78 bp to 2942 bp; peptide length: 955

Category: strong similarity to known protein

```

1 MPAVSKGDGM RGLAVFISDI RNCKSKEAEI KRINKELANI RSKFKGDKAL
51 DGYSKKKYVC KLLFIFLLGH DIDFGHMEAV NLLSSNKYTE KOIGYLFISV
101 LVNSNSELIR LINNAIKNDL ASRNPTFMCL ALHCIANVGS REMGEAFAAD
151 IPRLVAGDS MDSVKQSAAL CLLRLYKASP DLVPMGEWTA RVVHLLNDQH
201 MGVVTAAVSL ITCLCKKNPD DFKTCVSLAV SRLSRIVSSA STDLDQDYTY
251 FVPAPWLSVK LLRLQCYPY PEDAAVKGR LVECLETVLNK AQEPPKSKKV
301 QHSNAKNAIL FETISLIHY DSEPNLLVRA CNQLGQFLQH RETNLRYLAL
351 ESMCTLASSE FSHEAVKTHI DTVINALKTE RDVSVRQRAA DLYAMCDRS
401 NAKQIVSEML RYLETADYAI REEIVLKVAI LAEKYAVDYS WYVDITLNL
451 RIAGDYVSEE VWYRVLQIVT NRDDVQGYAA KTVFEALQAP ACHENMVKVG
501 GYLGEFGNL IAGDPRSSPP VQFSLHSHK HLCSVATRAL LLSTYIKFIN
551 LFPETKATIQ GVLRAQSQR NADVELQORA VEYLTLSVA STDVLATVLE
601 EMPFFPERES SILAKLKRKK GPGAGSALDD GRDPSSNDI NGGMEPTPST
651 VSTPSPSADL LGLRAAPPA APPASAGAGN LLVDVFDGPA AQPSLGPTPE
701 EAFSLSPGED IGPPPIPEADE LLNKFVCKNN GVLFENQLLO IGVKSEFRON
751 LGRMYLFYGN KTSVQFQNF PTVVHFGDLQ TQLAVQTKRV AAQVDGGAQV
801 QQVLNIECLR DFLTPPLSV RFRYGGAPQA LTLKLPVTIN KFFQPTMAA
851 QDFEQRWQL SLPQQAQKI FKANHPMDAE VTKAKLLGFG SALLDNVDPN
901 PENEVGAGII QTKALQVGCL LRLEFNAQAQ MYRLTLRTSK EPVSRHLCCL
951 LAQQF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_20h13, frame 3

PIR:B30111 alpha-adaptin C - mouse, N = 1, Score = 3990, P = 0

PIR:S11276 alpha-adaptin c - rat, N = 1, Score = 3987, P = 0

SWISSPROT:ADAC_RAT ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HA2/AP2 ADAPTIN ALPHA C SUBUNIT)., N = 1, Score = 3982, P = 0

SWISSPROT:ADAC MOUSE ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HA2/AP2 ADAPTIN ALPHA C SUBUNIT).., N = 1, Score = 3976, P = 0

TREMBL:AB020706_1 gene: "KIAA0899"; product: "KIAA0899 protein"; Homo sapiens mRNA for KIAA0899 protein, partial cds., N = 1, Score = 3932, P = 0

>PIR:B30111 alpha-adaptin C - mouse
Length = 938

HSPs:

Score = 3990 (598.6 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 787/955 (82%), Positives = 858/955 (89%)

```

Query:      1 MPAVSKGDMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60
             MPAVSKGDMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC
Sbjct:      1 MPAVSKGDMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60

Query:     61 KLLFIFLLGHDIIDFGHMEAVNLLSSNKYTEKQIGYLFISVLVNSNSELIRLINNAIKNDL 120
             KLLFIFLLGHDIIDFGHMEAVNLLSSN+YTEKQIGYLFISVLVNSNSELIRLINNAIKNDL
Sbjct:     61 KLLFIFLLGHDIIDFGHMEAVNLLSSNRYTEKQIGYLFISVLVNSNSELIRLINNAIKNDL 120

Query:    121 ASRNPTFMCLALHCIAVNGSREMGEAFAADIPRILVAGDSMDSVKQSAALCLLRLRYKASP 180
             ASRNPTFM LALHCIAVNGSREM EAFA +IP+ILVAGD+MDSVKQSAALCLLRLRY+ SP
Sbjct:    121 ASRNPTFMGLALHCIAVNGSREMAEAFAGEIPKILVAGDTMDSVKQSAALCLLRLRYTSP 180

Query:    181 DLVPMGEWTARVVHLLNDQHMGVVTAASVSLITCLCKKNPDDFKTCVSLAVSRLSRIVSSA 240
             DLVPMG+WT+RVVHLLNDQH+GVVTA SLIT L +KNP++FKT VSLAVSRLSRIV+SA
Sbjct:    181 DLVPMGDWTSRVVHLLNDQHLGVVTAATSLITTLAQKNPEEFKTSVSLAVSRLSRIVTSA 240

Query:    241 STDLDQDYTYFVPAPWLSVKLLRLLQCYPPPEDAAVKGRLVECLETVLNKAQEPKSKKV 300
             STDLDQDYTYFVPAPWLSVKLLRLLQCYPPP D AV+GRL ECLET+LNKAQEPKSKKV
Sbjct:    241 STDLDQDYTYFVPAPWLSVKLLRLLQCYPPP-DPAVRGRLTECLETILNKAQEPKSKKV 299

Query:    301 QHSNAKNAILFETISLIIHYDSEPNLLVRACNLQGQFLQHRETNRLRYLAESMCTLASSE 360
             QHSNAKNA+LFE ISLIIH+DSEPNLLVRACNLQGQFLQHRETNRLRYLAESMCTLASSE
Sbjct:    300 QHSNAKNAVLFEAISLIIHHDSEPNLLVRACNLQGQFLQHRETNRLRYLAESMCTLASSE 359

Query:    361 FSHEAVKTHIDTVINALKTERDVSVRQRAADLLYAMCDRSNAQIVSEMRLRYLETADYAI 420
             FSHEAVKTHI+TVINALKTERDVSVRQRA DLYAMCDRSNA+QIV+EML YLETADY+I
Sbjct:    360 FSHEAVKTHIETVINALKTERDVSVRQRAVDLLYAMCDRSNAQIVAEMLSYLETADYSI 419

Query:    421 REEIVLKVAILAEKYAVDYSWYVDITILNLIRIAGDYVSEEVWYRVQLQIVTNRDDVQGYAA 480
             REEIVLKVAILAEKYAVDY+WYVDITILNLIRIAGDYVSEEVWYRV+QIV NRDDVQGYAA
Sbjct:    420 REEIVLKVAILAEKYAVDYTWYVDITILNLIRIAGDYVSEEVWYRVQIVINRDDVQGYAA 479

Query:    481 KTVFEALQAPACHENMVKVGYYILGEFGNLIAGDPRSSPPVQFSLHLSKFLCSVATRAL 540
             KTVFEALQAPACHEN+VKVGYYILGEFGNLIAGDPRSSP +QF+LLHLSKFLCSV TRAL
Sbjct:    480 KTVFEALQAPACHENLVKVGYYILGEFGNLIAGDPRSSPLIQFNLLHLSKFLCSVPTRAL 539

Query:    541 LLSTYIKFINLFPETKATIQQVLRAGSQLRNADVELQQRAVEYLTLSVASTDVLATVLE 600
             LLSTYIKF+NLFPE KATI Q VLR+ SQL+NADVELQQRAVEYL LS+VASTD+LATVLE
Sbjct:    540 LLSTYIKFVNLFPEVKATI QDVLRSDSQLKNADVELQQRAVEYLRSTVASTDILATVLE 599

Query:    601 EMPFFPERESSILAKLRKKGPAGSALDDGRDPSSNDINGGMEPTP---STVSTPSPS 657
             EMPFFPERESSILAKLK+KKGP + L++ +R+ S D+NGG EP P S STPSPS
Sbjct:    600 EMPFFPERESSILAKLRKKGPSTVTDLEETKRERSI-DVNGGPEPVPASTSAASTPSPS 658

Query:    658 ADLLGLRAAPP-PAAPPASAGAGNLLVDVFDGPAAPSLGPTPEEAFLSPGPEDIGPPIP 716
             ADLLGL A PP P PP S+G G LLVDVF A+ ++ P L+PG ED
Sbjct:    659 ADLLGLGAVPPAPTGPSSGGG-LLVDVFDSDAS--AVAP-----LAPGSEDN----- 704

Query:    717 EADELLNKFVCKNNGVLFENQLLQIGVSKSEFRQNLGRMYLFYGNKTSVQFQNFSPTVVHP 776
             +FVCKNNGVLFENQLLQIG+KSEFRQNLGRM++FYGNKTS QF NF+PT++
Sbjct:    705 -----FARFVCKNNGVLFENQLLQIGLSEFRQNLGRMFFFYGNKTSQFLNFTPTLICA 759

Query:    777 GDLLQTQLAVQTKRVAAQVDGGAQVQVNLNIECLRDFTPLLSVRFYGGAPQALTCLKLP 836
             DLQT L +QTK V VDGAQVQV+NIEC+ DF P+L+++FRYGG Q +++KLP
Sbjct:    760 DDLQTNLNLQTKPVDPTVDGGAQVQVNLNIECISDFTEAPVLNIQFRYGGTFQNVSVKLP 819

Query:    837 VTINKFFQPTEMAQDFFQRWKQLSLPQQEAQKIFKANHPMDAEVTKAKLLGFGSALLDN 896
             +T+NKFFQPTEMA+QDFFQRWKQLS PQQE Q IFKA HPMD E+TKAK++GFGSALL+
Sbjct:    820 ITLNKFFQPTEMAQDFFQRWKQLSNPQQEVQNIKAKHPMDTEITKAKIIGFGSALLEE 879

Query:    897 VDPNPENFVGAGIIQTALQVGCLLRLLEPNAAQMYRLTLRTSKEPVSRHLCELLAQGF 955
             VDPNP NFVGAGII TK Q+GCLLRLEPN QAQMYRLTLRTSK+ VS+ LCELL++QF

```

Report for DKFZphutel 20h13.3

[illegible]


```

PRD      eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
SEQ      LLNKFVCKNNGVLFFENQLLQIGVKSEFRQNLGRMYLFYGNKTSVQFQNFSPTVVHPGDLQ
SEG      .....
PRD      ceeeeeeeeccccchhhhhhhcchhhhhccccceeeeeccccccccccccceeeeeccccchh

SEQ      TQLAVQTKRVAAQVDGGAQVQVNLNIECLRDFTLPPLLSVRFRYGGAPQALTCLKLPVTIN
SEG      .....xxxxxxxxxxxxx.....
PRD      hhhhhhhccccccccchhhhhhhhhccccccccceeeeecccccccccccccccccc

SEQ      KFFQPTEMAAQDFQRWKQLSLPQQEAQKIFKANHPMDAEVTKAKLLGFGSALLDNVDPN
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhcchhhhhhhhhccccchhhhhhhhhccccceeeeecccc

SEQ      PENFVGAGIIQTKALQVGCLLRLEPNAQAQMYRLTLRTSKEPVSRHLCELLAQQF
SEG      .....
PRD      cccececececececececececcccchhhhhhhhhccccchhhhhhhhhcccc

```

Prosites for DKFZphute1_20h13.3

PS00001	760->764	ASN_GLYCOSYLATION	PDOC00001
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	163->166	PKC_PHOSPHO_SITE	PDOC00005
PS00005	189->192	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	297->300	PKC_PHOSPHO_SITE	PDOC00005
PS00005	379->382	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	470->473	PKC_PHOSPHO_SITE	PDOC00005
PS00005	787->790	PKC_PHOSPHO_SITE	PDOC00005
PS00005	819->822	PKC_PHOSPHO_SITE	PDOC00005
PS00005	832->835	PKC_PHOSPHO_SITE	PDOC00005
PS00005	935->938	PKC_PHOSPHO_SITE	PDOC00005
PS00005	938->941	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	368->372	CK2_PHOSPHO_SITE	PDOC00006
PS00006	379->383	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	482->486	CK2_PHOSPHO_SITE	PDOC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDOC00006
PS00006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS00006	636->640	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00006	938->942	CK2_PHOSPHO_SITE	PDOC00006
PS00007	388->395	TYR_PHOSPHO_SITE	PDOC00007
PS00007	411->419	TYR_PHOSPHO_SITE	PDOC00007
PS00007	434->443	TYR_PHOSPHO_SITE	PDOC00007
PS00008	202->208	MYRISTYL	PDOC00008
PS00008	508->514	MYRISTYL	PDOC00008
PS00008	561->567	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00008	759->765	MYRISTYL	PDOC00008
PS00008	826->832	MYRISTYL	PDOC00008
PS00008	908->914	MYRISTYL	PDOC00008
PS00009	630->634	AMIDATION	PDOC00009
PS00290	127->134	IG_MHC	PDOC00262

(No Pfam data available for DKFZphute1_20h13.3)

DKFZphut1_20ml1

group: cell cycle

DKFZphut1_20ml1 encodes a novel 225 amino acid protein with similarity to yeast sds22 and protein phosphatase-1 regulatory subunits.

sds22 is a regulatory polypeptide of protein phosphatase-1 that is required for the completion of mitosis in both fission and budding yeast. The novel protein seems to be a new regulator protein for protein phosphatase-1.

The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

similarity to suppressor protein sds22

complete cDNA, complete cds, EST hits
localisation? only a part of the STS matches

Sequenced by AGOWA

Locus: /map="17"?

Insert length: 5822 bp

Poly A stretch at pos. 5803, polyadenylation signal at pos. 5786

```
1 GGGCGCTTGG TTCCCCAGCA ACCGGGAGAC GCGTCTGCTG CGTGGAAACCG
51 CCGAGTTCCTC AGCGCTTGAG AAGGAAAATT CTGGATCTGT TATCTGTGAG
101 GAGGCCACTC CGTTGACAGT TGTGTAAAAC TCTGCTGCTT TCCCCAGCTC
151 CAACCTCTCT GGTCTTCAAC AACACTATCA TCAGGGAAAA CGTGGGGGAA
201 GATGAACCAAG CCGTGCAACT CGATGGAGCC GAGGGTGTAT GACGATGACA
251 TGCTCAAGCT GGCCGTCGGG GACCAGGGCC CCCAGGAGGA GGCCGGGCAG
301 CTGGGCAAGC AGGAGGGCAT CCTCTCAAG GATGTCTGT CCCTGCAGCT
351 GGACTTTCGG AACATCCTCC GCATAGACAA CCTCTGGCAG TTTGAGAAGT
401 TGAGGAAGCT GCAGCTGGAC AATAACATCA TTGAGAAGAT CGAGGGCCTG
451 GAGAACTCTG CACACCTGGT CTGGCTGGAT CTGTCTTTCA ACAACATTGA
501 GACCATCGAG GGGCTGGACA CACTGGTGAA CCTGGAGGAC CTGAGCTTGT
551 TCAACAACCG GATCTCCAAG ATCGACTCCC TGGACGCCCT CGTCAAGCTG
601 CAGGTGTGTT CGCTGGGCAA CAACCGGATT GACAACATGA TGAACATCAT
651 CTACCTCCGG CGGTTCAAGT GCCTGCGGAC GCTCAGCCTC TCTAGGAACC
701 CTATCTCTGA GGCAGAGGAT TACAAGATGT TCATCTGTGC CTACCTTCCT
751 GACCTCATGT ACCTGGACTA CCGGCGCATT GATGACCACA CAGCAAGTGT
801 CTCCCTCTCA GTCTCCAGC CTTGTGAGAC AGATTCTCTA AGCCCCAGG
851 TTTCTTGGAA AAGGGGCATT GAAGAGTAGC TTCCCTGCC CACAACCTAGG
901 AGAGAAAGGG CAGCTCCCTC TTCCTAATCC CTTTACCTGA CTCTGTCAGA
951 GTGATTCCAG CAGCACCTT GTAAGTACTG TTTTGTGTGC GTTCCAGGG
1001 GCCAGGCCCTC TTCCACACAC TGTCCAGGG CCACCTCACA GCCATCCTGC
1051 ACTGTCTAGT TTTCCAGATG AAGAAGCTGA GGAGGGCTGG GAGCAGTGGC
1101 TACAGCCTGT AATCCCAGCA CTTTGAGAGG CTGAGGCGGG AGGATCGCTT
1151 GAGCCAAGGA GTTCAAGACC AGCCTGGGCA ACATAGGGAG ACCCCATCTC
1201 TACAGAAACT ACCAAAATTA GCCAGGTGTG GTGGCACACA CCAGTAATCC
1251 TGGCTACTCA CAAGGCCGAG GTAGAAGAAT CGCTTGAGAC TAGGAGTTTG
1301 AGGCTGCACT GAACTAAGAA GATGCCATTG CACTCCAGCC TGGGCAACAG
1351 AGTGAAAAAA TTAATAAATT AGAAAAGAAA AGAAGTTGAG GAGGCCCAAG
1401 GAGGCCAAGC AGCCAGGATC ACTGGCTCAA GGCCAAGCCA GGATTCAACC
1451 TAAGTTGGTG TCATCCCAGG AGCAATATTA ACAGCTGAGC TCCAGAGGGA
1501 ACCAGGCCAT CAGAGGCTCA GGCTTGGCTC TCAGGGGCAG AGTCAGGGCT
1551 GGAGGTAGAG ACCTGAGTGT CATCTGAGGA TTGCCAATTG GCAGTAGTTG
1601 AAGCCATGGT ACAGGTGGGA TCACCTGGGG CACATGGAGT GAGCTGGGGG
1651 ACGGGGACTA AGTTCTAGAG GTGCCAGCAT TCCTGGCCAG GTACAGGGGG
1701 ATGAGCCAGT GCGGTGGAGA GAGCCAAGGG CCAGACCCTC GTGACCAGCC
1751 CTATGGCCTC ACTCTACCTC TGTCTGTGTT TCCTCCTTCC CTAAGAGAGG
1801 GCCAGAAGGC CTGCTGAGGG CTGTTGGGAG TGAGAGAGCA AGTCCTCTGT
1851 GGAGAACACC CAGTCTGGGG CGAGGGGAGC GCTCCATTGC TGTGGCTCCT
1901 GCCCTGAGAG TGGCCCGGG AACCCAGCC TGCCACGCTG CCTTCGCTC
1951 TCCCTGGTCT TTCCCTGATT TCCCTGCGCT CACAAAAACC TGGTGAGGGT
2001 CATCAGGAGA TGGGCATTCT CATCCACGAG ACCTCATGGC TTTCACAGCC
2051 TTCATGCGAG CCCCTGTGCA ACACCCCTGC CCATGCGCGG GAGGCTGCAG
2101 CATGGCAGAG GCGGCATGGC AGAGGCGGTG TGGCTCGGAG GAACCTCTGG
2151 TAACAATGCC ACTCCCGTTC CTTGGTCAGA AAAAGCTTGC GGAGGCTAAG
2201 CACCAGTACA GCATCGACGA GCTGAAGCAC CAGGAGAACG TGATGCAGGC
2251 CCAGCTGGAG GACGAGCAGG CGCAGCGGGA GGAGCTAGAG AAGCACAAGA
2301 CTGCGTTTGT GGAACACCTG AATGGCTCCT TCCTGTTTGA CAGCATGTAC
2351 GCTGAGGACT CAGAGGGCAA CAATCTGTCC TACCTGCCTG GTGTCGGTGA
2401 GCTCCTTGAG ACCTACAAGG ACAAGTTTGT CATCATCTGC GTGAATATTT
2451 TTGAGTATGG CCTGAAACAG CAGGAGAAGC GGAAAACAGA GCTTGACACC
2501 TTCAGTGAAT GTGTCCGTGA GGCCATCCAG GAAAACAGG AGCAGGGCAA
```

```

2551 ACGCAAGATT GCCAAATTCG AGGAGAAGCA CTTGTGCGAGT TTAAGTGCCA
2601 TTCGAGAGGA GTTGGAACTG CCCAACATTG AGAAGATGAT CCTAGAATGC
2651 AGTGCTGACA TCAGTGAGTT GTTCGATGCG CTCATGACGC TGGAGATGCA
2701 GCTGGTGGAG CAGCTGGAGG TAAGGCTGGG CCCTGGGCAC AAGTGCCAGA
2751 ATCTGGCGAT GCAGCTGCAC ATCCATAGGT GAACTGTAGC CTTATGGGCG
2801 ACGCCTCTGC TGGAAACGTC CAGCACGACT CAGCGTGGCA GGCTGTAGCT
2851 TTCTTGCTCA TCAGTCCTGT TTGCTTTTAT TACATTTTAA TCATTTACAT
2901 TGGAAAGTAT TCTTGTGGAA AATGAGAGGT GAGCTCATTG TTCTGAAATG
2951 GTCCCCCTAT CCTGGGAAGTC AGTGGGGAGA GGTTTTGTAT TAGACCCCTG
3001 GAGCTATCCG GGTACTCTAA AGGCAAAGCG CACCCCCACT TGGGGACCAA
3051 ACAAAGACCC CTCCGCATTG CAGCCTGCAG TTGCCCTTC TCAGGTGACG
3101 TGAGGAGGCT GCAACTCAGC ACTAAGTAGT GAAAAAGAAA AGCGCCGCTG
3151 TCTGAAATTC ATTAGCAGCC AGAGTATGTG TTACAAGGCA GCGGAGGCTG
3201 GGAGTCTGAA GTGGTGTGAT GAATTGAACC TCATCGGATG CTGCTGTGGC
3251 TGGGCCAAGT GATAGCACCT AATCAATTCC TCACACGTCA AGTGACACCT
3301 CAGACATGGG ATAGATTTC CCATCACATC ACAGGGCAGG TGCTCCCTCC
3351 CTGCTGGAGA GCACAGGCAC TGCAGAAGCA GCGCACAGTG CCAGGGGCGA
3401 GTGAGGCAGC AGCTCCCAGC CTTTTCAGGC ACGGAGATTG CCTTTCAACA
3451 TCCAAACATT TCCCAGAAC CATGTGCCAT CCTACTGTGA TTACTGGTGG
3501 CCAGAAAGCC ACAAGCGCAA TCATGCTTTT CAATGACCCT ATTTTATTTC
3551 ACGAGAACAG CACATACATG TGTTTGAAAA TTATGTGAGG TGCTCACTCT
3601 GCAGACAGTA CTCACATTCC TATAGATTCC ACCCCTGCCC ACCTTGCAGC
3651 CCCTGGAGTC TATAGCAGAT GGGAGTGGGG CACTCCGAGA GTGGCAGGCG
3701 TGGAGATCAC ATCTTCCATT GTTCCTTCAA TCAACACTAA CTCCCATTG
3751 GGCCTTAGGT GCCTTGCTAA GCACCACAAA ACAGCAACTA ACTGAAAGAG
3801 ATCTGGAGTG CCAGCCCGCT CCTACTGAGG GCCTCCTCTC TGTCAGGCAC
3851 CTTGCAAAGC ATTTTGTGTG AAGTGACTCA TTTAACTTCA CCACAACGCC
3901 ACAACGCAGG GATTATGCAG GTAACCTATT TCCCAGATGA GGAAGATAAG
3951 GCCCCAAGGAG GTGAAATGCC TTTCCCAGAG TTACACAGAG TGCTGGAGCT
4001 GGAATACTAG ACCCAGGCAG TCTAGCTCTT AACAGCTCAC TCCACTGTTT
4051 CCCTGGAGGT GATGCACAGA TGCTACTGGG AAACCCAAAG GAGAGGGGGT
4101 TGGCTGTGTG TGTGTGTGTT GGGCAGGCAG GTAAGGGGAG TAAGACCAGG
4151 ACAAGTGTTC CTGGCAAAGT TCCGGTGACA GCATTAAACA TTCAGATGGT
4201 GAGGGAGTTA ATATGGTTGG AGAACAACAA CTTTAGAGAG AGCAGAGGGG
4251 TCAGTTCACA ACCATCTGCT CAGGAGGGTC AAGATGGGTG GTCTTTATGC
4301 TGAAGGTCTG TGATTAGAGG AGCTGGTTGC TAAATTTTGA GGAGTACCTT
4351 TTGCTCTGTG CTGGACATCT AAATATGCAT GTTAACCTGTG TTCTTTAACA
4401 TTTCCAGGAG ACTATAAACA TGTTTGAAAG GAACATTGTT GACATGGTAG
4451 GACTGTTTAT CGAAAAATGTC CAAAGCCTAT ATCCTTTCTG TGATGACCTT
4501 CCCCATTGGG AGGTGCTACA GAGCCCTCGG GCTTGTCCCG GCCTCTGGAC
4551 AAAAGAAATG TCCACAGGGT CTGAGGAGGT TTCCCAGACC TCAGAACAAT
4601 GATGGCCTCG TTAGAGCTGT GGTTTGGATG CCCAGAGGGA CAACATCCAA
4651 ACTGTTTGCA GTAGGCTCCC AGCATGATTG TTCTCATATG AGTGATGTTT
4701 ACTAGGAAAT GACGCCCCCT GTGTTGCAGG CAAGCACACT CTGGGGTTGA
4751 GGCAACCCCC ACGTGGAAGA CACTATAAGG AGTACATCAG GTGAAATGTT
4801 AGGGTGAGGA GCCAACATCG GAGCATGGCC AACCTTCTT CCACCCGAAC
4851 TCAGGGCACT CCACATGGGG CAAACTGCTG TGCTCCAGCT AGCAGCAGCC
4901 CTGTGGTCTT GCCCTCCTGG GGCTCACAGT CCCTCAGGGA GACAAGTTGT
4951 AGAGGCAACA AGTGGTGCCA AATGCACAGG GTGAGAAGCA GTTAACCCAG
5001 AGGCCAGGAG CCTCCATGCA GGAGGAGAG AAGAGTGTGA TGGCAGGGGC
5051 CGAGGGTCCG TCCGAGGTGT GGGGCAGGGG CAGGGAGTCG AGGAAGGGCC
5101 AGGGTTCCGA GCTTGTGAGT GGACGGTGCT GCCAGCCAGA ATTTCCGAGC
5151 TCGCCTTGGG CCCTTAAAGT CTGTCTCCCG CCGTCTGAGA GCATCAGGGA
5201 CGCGCCGGGC CTGCTCCTCC CGGGCCTTTG CTTAACTCGG GGCTGCACGA
5251 TGGCTCAGTG CCGGGACCTG GAGAATCACC ACCACGAGAA GCTCCTGGAG
5301 ATCTCTATCA GCACCTTGGA GAAGATTGTC GAGGGCGACC TGGACGAGGA
5351 CCTGCCTAAC GACCTCGCGG CGCTTTTGTG CGATAAAGAT ACGATTGTTA
5401 ATGCTGTCCG GGCATCGCAC GACATCCACC TCCTGAAGAT TGACAATCGA
5451 GAAGATGAGC TGGTGACCAG AATCAACTCT TGGTGTACAC GTTTAATAGA
5501 CAGGATTCAC AAGGATGAGA TCATGAGGAA CCGCAAGCGC GTGAAGGAGA
5551 TCAATCAGTA CATCGACCAC ATGCAGAGCG AACTGGACAA CCTGGAATGT
5601 GGCAGCATCC TAGACTAGAT GAATGTACGC CACAGGAGCT TCTTCAAAAC
5651 ATAGCACCAG CCCCAGCCAG GAGAAGGAAG TGCACACGCC TCACCCGCAC
5701 CTCTAGAGAG TTGCTGGGCA TCTCTCAACC GCGATCCCCA ACACCATTTCT
5751 TCCCCCACC CTGGAAAAAC TTCCAAAAGT AGAGAAAATA AAGGACTCAT
5801 TTCACAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HS1292248 from database EMBL:

human STS SHGC-53917.

Score = 874, P = 3.3e-33, identities = 180/185

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 876 bp; peptide length: 225
Category: similarity to known protein

1 MNQPCNSMEP RVMDDMLKL AVGDQGPQEE AGQLAKQEGI LFKDVLSLQL
51 DFRNLRIDN LWQFENLRKL QLDNNIEKI EGLENLAHLV WDLFSFNNIE
101 TIEGLDTLVN LEDLSLFNNR ISKIDSLDAL VKLQVLSLGN NRIDNMMNII
151 YLRRFKCLRT LSLSRNPISE AEDYKMFICA YLPDLMYLDY RRIDDHTASV
201 SLSVSQPCET DSSSPQVSWK RGIEE

BLASTP hits

Entry S68209 from database PIR:
sds22 protein homolog - human >TREMBL:HSSDS22MR_1 gene: "sds22";
product: "yeast sds22 homolog"; H.sapiens sds22-like mRNA
Score = 234, P = 1.2e-19, identities = 61/143, positives = 93/143

Entry A38439 from database PIR:
suppressor protein sds22(+) - fission yeast (Schizosaccharomyces pombe)
>TREMBL:SPSDS22_1 gene: "sds22+"; S.pombe sds22+ gene, complete cds.
Score = 208, P = 5.6e-17, identities = 52/127, positives = 71/127

Entry S43988 from database PIR:
protein suppressor sds22 - fission yeast (Schizosaccharomyces pombe)
>SWISSPROT:SD22_SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT
SDS22. >TREMBL:SPAC4A8_12 gene: "sds22"; product: "phosphatases pp1
regulatory subunit"; S.pombe chromosome I cosmid c4A8.
Score = 208, P = 8.5e-17, identities = 52/127, positives = 71/127

Entry CEK10D2_5 from database TREMBL:
gene: "K10D2.1"; Caenorhabditis elegans cosmid K10D2.
Score = 214, P = 3.6e-16, identities = 50/125, positives = 75/125

Alert BLASTP hits for DKFZphut1_20ml1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_20ml1, frame 1

Report for DKFZphut1_20ml1.1

[LENGTH] 225
[MW] 25955.87
[pI] 4.63
[HOMOL] PIR:S68209 sds22 protein homolog - human 1e-18
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL193c] 2e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL193c] 2e-11
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 2e-11
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YOR373w] 2e-06
[FUNCAT] 01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae,
YJL005w] 3e-05
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-05
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-05
[FUNCAT] 10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-05
[FUNCAT] 04.07 rna transport [S. cerevisiae, YPL169c] 9e-04
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR065w] 9e-04
[EC] 4.6.1.1 Adenylate cyclase 2e-06
[PIRKW] nucleus 5e-16
[PIRKW] duplication 2e-06
[PIRKW] tandem repeat 2e-06
[PIRKW] cAMP biosynthesis 2e-06
[PIRKW] glycoprotein 2e-06
[PIRKW] phosphorus-oxygen lyase 2e-06
[SUPFAM] leucine-rich alpha-2-glycoprotein repeat homology 5e-16
[SUPFAM] fibromodulin 3e-07
[SUPFAM] yeast adenylate cyclase catalytic domain homology 2e-06
[SUPFAM] yeast adenylate cyclase 2e-06
[PROSITE] CK2_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 1

{KW} All_Alpha

```

SEQ  MNQPCNSMEPRVMDDMLKLAVGDQGPQEEAGQLAKQEGILFKDVLSQLDFRNILRIDN
PRD  cccccccccccccchhhhhccccchhhhhhhhhchhhhhhhhhcccccccccc

SEQ  LWQFENLRKLQLDNNIIEKIEGLENLAHLVWLDLSFNNIETIEGLDTLVNLEDLSLFNNR
PRD  hhhhhhhhhhhccccccccccccchhhhhhhccccccccccccchhhhhhhhhcccccc

SEQ  ISKIDSLDALVKLVLSLGNRRIDNMMNIYLRRFKCLRTLSSLRNPISEAEDYKMFICA
PRD  cccchhhhhhhhhhhccccccccccccchhhhhhhhhccccccccchhhhhhhhh

SEQ  YLPDLMYLDYRRIDDHTASVLSVSQPCETDSSSPQVSWKRGIEE
PRD  hhccccccccccccchhhhhhhcccccccccccccccccccccc

```

Prosites for DKFZphut1_20ml1.1

PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	169->173	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphut1_20ml1.1)

DKFZphut1_20m24

group: metabolism

DKFZphut1_20m24 encodes a novel 611 amino acid protein with similarity to a hypothetical *C.elegans* protein and to yeast Alg9 protein.

This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2.

The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

strong similarity to *S.cerevisiae* Alg9p

complete cDNA, complete cds, potential start at Bp 23, few EST hits
Alg9 is involved in the assembly of the core oligosaccharide
Glc3Man9GlcNAc2

HSAC381 corresponding genomic DNA (2 exons)
HSB8954 corresponding genomic DNA (1 exon)

Sequenced by AGOWA

Locus: /map="11"

Insert length: 1986 bp

Poly A stretch at pos. 1966, polyadenylation signal at pos. 1949

```
1 TTCTTTTTC CCCAGGCTTG CCATGGCTAG TCGAGGGGCT CGGCAGCGCC
51 TGAAGGGCAG CGGGGCCAGC AGTGGGGATA CGGCCCCGGC TCGCGACAAG
101 CTGCGGGAGC TGCTGGGCAG CCGAGAGGCG GCGCGCGCGG AGCACC GGAC
151 CGAGTTATCT GGAACAAAG CAGGACAAGT CTGGGCACCT GAAGGATCTA
201 CTGCTTTCAA GTGCTGCTT TCAGCAAGGT TATGTGCTGC TCTCTGAGC
251 AACATCTCTG ACTGTGATGA AACATTCAAC TACTGGGAGC CAACACACTA
301 CCTCATCTAT GGGGAAGGGT TTCAGACTTG GGAATATTCC CCAGCATATG
351 CCATTGCTCT CTATGCTTAC CTGTTGCTTC ATGCCTGGCC AGCTGCATTT
401 CATGCAAGAA TTCTACAAAC TAATAAGATT CTGTGTTTT ACTTTTGGCG
451 ATGTCTTCTG GCTTTTGTGA GCTGTATTG TGAACTTAC TTTTACAAGG
501 CTGTGTGCAA GAAGTTTGGG TTGCACGTGA GTCGAATGAT GCTAGCCTTC
551 TTGGTTCTCA GCACTGGCAT GTTTTGCTCA TCATCAGCAT TCCTTCCTAG
601 TAGCTTCTGT ATGTACACTA CGTTGATAGC CATGACTGGA TGGTATATGG
651 ACAAGACTTC CATTGCTGTG CTGGGAGTAG CAGCTGGGGC TATCTTAGGC
701 TGGCCATTCA GTGCAGCTCT TGGTTTACCC ATTGCCTTTG ATTTGCTGGT
751 CATGAAACAC AGGTGGAAGA GTTCTTTTCA TTGGTCGCTG ATGGCCCTCA
801 TACTATTCTT GGTGCCTGTG GTGGTCATTG ACAGCTACTA TTATGGGAAG
851 TTGGTGATTG CACCACTCAA CATGTTTTGG TATAATGTCT TTACTCCTCA
901 TGGACCTGAT CTTTATGGTA CAGAACCCTG GTATTTCTAT TTAATTAATG
951 GATTTTCTGAA TTCAATGTA GCCTTTGCTT TGGCTCTCCT AGTCCTACCA
1001 CTGACTTCTC TTATGGAATA CCTGCTGCAG AGATTTCATG TTCAGAATTT
1051 AGGCCACCCC TATTGGCTTA CCTTGGCTCC AATGTATATT TGGTTTATAA
1101 TTTTCTTCAT CCAGCCTCAC AAAGAGGAGA GATTTCTTTT CCCTGTGTAT
1151 CCACTTATAT GTCTCTGTGG CGCTGTGGCT CTCTCTGCAC TTCAGAAATG
1201 TTACCACTTT GTGTTTCAAC GATATCGCCT GGAGCACTAT ACTGTGACAT
1251 CGAATTGGCT GGCATTAGGA ACTGTCTTCC TGTGTGGGCT CTTGTCAATT
1301 TCTCGCTCTG TGGCACTGTT CAGAGGATAT CACGGGCCCC TTGATTGTGA
1351 TCCAGAATTT TACCGAATTG CTACAGACCC AACCATCCAC ACTGTCCCAG
1401 AAGGCAGACC TGTGAATGTC TGTGTGGGAA AAGAGTGGTA TCGATTTCCC
1451 AGCAGCTTCC TTCTTCTGTA CAATTGGCAG CTTCAAGTCA TTCCATCAGA
1501 GTTCAGAGGT CAGTTACCAA AACCTTTTGC AGAAGGACCT CTGGCCACCC
1551 GGATTGTTCC TACTGACATG AATGACCAGA ATCTAGAAGA GCCATCCAGA
1601 TATATTGATA TCAGTAAATG CCATTATTTA GTGGATTGG ACACCATGAG
1651 AGAAACACCC CGGGAGCCAA AATATTTCAT CAATAAGAA GAATGGATCA
1701 GCTTGGCCTA TAGACCATTG CTTGATGCTT CTAGATCTTC AAAGCTGCTG
1751 CGGGCATTTCT ATGTCCCTT CCTGTGAGAT CAGTATACAG TGTACGTAAA
1801 CTACACCATC CTCAACCCCC GGAAGCAAAA GCAATCAGG AAGAAAAGTG
1851 GAGGTTAGCA ACACACCTGT GGGCCCAAAG GACAACCATC TTGTTAACTA
1901 TTGATTCCAG TGACCTGACT CCCTGCAAGT CATCGCCTGT AACATTTGTA
1951 ATAAAGGTCT TCTGACATGA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HSAC381 from database EMBL:
Homo sapiens chromosome 11 pac pDJ159ol, complete sequence.
Length = 42,771

Entry HSB8954 from database EMBL:

cSRL-50A3-u cSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone cSRL-50A3.
Length = 601

Medline entries

96293493:
Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of *Saccharomyces cerevisiae*: identification of the ALG9 gene encoding a putative mannosyl transferase.

Peptide information for frame 2

ORF from 23 bp to 1855 bp; peptide length: 611
Category: strong similarity to known protein

```

1 MASRGARQRL KGSGASSGDT APAADKLREL LGSREAGGAE HRTELSGNKA
51 GQVWAPEGST AFKCLLSARL CAALLSNISD CDETFNYWEP THYLIYGEF
101 QTWEYSPAYA IRSYAYLLH AWPAAFHARI LQTNKILVY FLRCLLAFVS
151 CICELYFYKA VCKKFGLHVS RMMLAFLVLS TGMFCSSAF LPSSFCMYTT
201 LIAMTGWYMD KTSIAVLGVA AGAILGWPF AALGLPIAFD LLVMKHRWKS
251 FFHWSLMALI LFLVPVVVID SYYYGKLVI PLNIVLYNVF TPHGPDLYGT
301 EPWYFYLING FLNFNVAFAL ALLVLPLTSL MEYLLQRFHV QNLGHPYWLT
351 LAPMYIWFII FFIQPHKEER FLFPVYPLIC LCGAVALSAL QKCYHFVFQR
401 YRLEHYTVTS NWLALGTVFL FGLLSFSRSV ALFRGYHGPL DLYPEFYRIA
451 TBPTIHTVPE GRPVNVCVGK EWYRFPSSFL LPDNWQLQFI PSEFRGQLPK
501 PFAEGPLATR IVPTDMNDQN LEEPSRYIDI SKCHYLVOLD TMRETPREP
551 YSSNKEEWS LAYRPFLDAS RSSKLLRAFV VPFLSDQYTV YVNYTILKPR
601 KAKQIRKKS G

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_20m24, frame 2

SWISSPROT:YTH3 CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

SWISSPROT:YTH3 CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

>SWISSPROT:YTH3 CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II.

Length = 653

HSPs:

Score = 957 (143.6 bits), Expect = 2.7e-96, P = 2.7e-96
Identities = 206/514 (40%), Positives = 296/514 (57%)

```

Query: 48 NKAGQVWAPEGSTAFKCLLSARLCAALLSNISDCDETFNYWEPHTHYLIYGEFQTWEYSP 107
      N   W   + FK LLS R+ A+ I+DCDE +NYWEP H +YGEFQTWEYSP
Sbjct: 43 NNPDNDWPFSFGSVFKMLLSIRISGAIWGIINDCDEVYNYWEPLHLFLYGEFQTWEYSP 102

Query: 108 AYAIRSYAYLLHAWPAAFHARILQTNKILVYFLRCLLAFVSCICELYFYKAVCKKFGL 167
      YAIRSY Y+ LH PA+ A + KI+VF +R + + E Y + A+CKK +
Sbjct: 103 VYAIRSYFYIYLHYIPASLFANLFGDTKIVVFTLIRLTIGLFCLLGEYYAFDAICKKINI 162

Query: 168 HVSRMMLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGW 227
      R + F + S+GMF +S+AF+PSSFCM T + + + + + VA ++GW
Sbjct: 163 ATGRFFILFSIFSSGMFLASTAFVPSSFCMAITFYILGAYLNENWTAGIFCVAFSTMVGW 222

Query: 228 PFSAALGLPIAFDILLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLY 287

```

Sbjct: 223 PFSFA LGLPI D+L++K F SL+ + V+ DS+Y+GK V+APLNI LY
 Query: 288 NVFTPHGPDLYGTEPWYFYLINGFLNFNVAFA LALLVLP L TSLMEYLLQRFHVQNLGHPY 347
 NV + GP LYG EP FY+ N F N+N+ A PL+ + Y + + Q+
 Sbjct: 283 NVVSGPGPSLYGEEPLSFYIKNLFNNWNIVIFAAPFGFPLS--LAYFTKVWMSQDRNVAL 340
 Query: 348 WLTAPMYI-----WFIFFIQPHKEERFLFPVYPLICGVALSALQKCYHFVQR 400
 + AP+ + W +IF Q HKEERFLFP+Y P I A+AL A + ++
 Sbjct: 341 YQRFAPILLAVTTAAWLLIFGSAHKEERFLFPIYPIAFFAALALDATNR---LCLKK 397
 Query: 401 YRLEHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPE 460
 ++ N L++ + F +LS SR+ ++ Y +++Y T+ T +
 Sbjct: 398 LGMD-----NLSILFILCFAILSASRTYSIHNNYGSHEIYRSLNAELTNRT-NFKNF 450
 Query: 461 GRPVNVCGKWEYRFPSSFLPDNW-----QLQFIPSEFRGQLPKPFAEGPL---ATRI 511
 P+ VCVGKEW+RFPSSF +P +++FI SEFRG LPKPF + TR
 Sbjct: 451 HDPTRVCVGKEWHRFPSSFFIPQTVSDGKKVEMRFIQSEFRGLLPKPLKSDKLVEVTRH 510
 Query: 512 VPTDMNDQNLEEPSRYIDISKCHYLVDLDTMRETPREPKYSSNKEEW 558
 +PT+MN+ N EE SRY+D+ C Y+VD+D M ++ REP + ++ +
 Sbjct: 511 IPTMNNLNQEEISRYVDLDSCDYVVDVD-MPQSDREPDFRKMQRNY 556

Pedant information for DKFZphut1_20m24, frame 2

Report for DKFZphut1_20m24.2

[LENGTH] 611
 [MW] 69863.78
 [pI] 8.91
 [HOMOL] SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II. 2e-93
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNL219c] 4e-69
 [FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YNL219c] 4e-69
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YNL219c] 4e-69
 [PIRKW] glycosyltransferase 9e-68
 [PIRKW] transmembrane protein 9e-68
 [PIRKW] hexosyltransferase 9e-68
 [PROSITE] MYRISTYL 9
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 7
 [PROSITE] PKC_PHOSPHO_SITE 6
 [PROSITE] ASN_GLYCOSYLATION 2
 [KW] TRANSMEMBRANE 7
 [KW] LOW_COMPLEXITY 6.71 %

SEQ MASRGARQRLKSGSGASSGDTAPAADKLRELLGSREAGGAEHRTLSGNKAGQVWAPEGST
 SEG
 PRD cccchhhhhhhccccccccccccchhhhhhhhhccccccccccccccccccccccccccccch
 MEMMMMMMM

SEQ AFKCLLSARLCAALLSNISDCDETFNYWEPHYLIYEGGFQTWEYSPAYAIRSYAYLLH
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhhhhhhhhc
 MEM MM

SEQ AWPAAFHARILQTNKILVFYFLRCLLAFVSCICELYFYKAVCKKFGHLVSRMMLAFLVLVS
 SEG
 PRD cchhhhhhhhhhhcc
 MEM MM

SEQ TGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD
 SEG
 PRD cccccccccccccchhhhhhhhhhhcc
 MEMMMMMMMMMMMMMMMMM

SEQ LLVMKHRWKSFFHWSLMALILFLVPVVVIDSYGKLVIAPLNIVLYNVFTPHGPDLYGT
 SEG
 PRD hhh
 MEM MMMMMMM.MMM

SEQ EPWYFYLINGFLNFNVAFA LALLVLP L TSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII
 SEG
 PRD cccccccccccccchhh
 MEMMM


```

SEQ  FFIQPHKEERFLFPVYPLICLCGAVALSALQKCYHFVFQRYRLEHYTVTSNWLALGTVFL
SEG  .....
PRD  hhccccchhhhhccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccchhhhhhhhe
MEM  .....MMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMM.....

SEQ  FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVGKEWYRFPSSFL
SEG  .....
PRD  eehhhhhhhheeeccccccccccccceeeccccccccceeecccccccccccccccccc
MEM  .....

SEQ  LPDNWQLQFIPSEFRGQLPKPFAEGPLATRIVPTDMNDQNLEEPSRYIDISKCHYLVOLD
SEG  .....
PRD  cccccceeeccccccccccccccccceeeccccccccccccccccceeeceeeceeecc
MEM  .....

SEQ  TMRETPREPKYSSNKEEWISLAYRPFLDASRSSKLLRAFYVPFLSDQYTVYVNYTILKPR
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhheeeeeeecceeeeeeecccc
MEM  .....

SEQ  KAKQIRKSGG
SEG  .....
PRD  hhhhhcccccc
MEM  .....

```

Prosites for DKFZphut1_20m24.2

PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	593->597	ASN_GLYCOSYLATION	PDOC00001
PS00004	606->610	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	67->70	PKC_PHOSPHO_SITE	PDOC00005
PS00005	133->136	PKC_PHOSPHO_SITE	PDOC00005
PS00005	541->544	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	553->556	PKC_PHOSPHO_SITE	PDOC00005
PS00005	572->575	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	329->333	CK2_PHOSPHO_SITE	PDOC00006
PS00006	457->461	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00006	545->549	CK2_PHOSPHO_SITE	PDOC00006
PS00006	553->557	CK2_PHOSPHO_SITE	PDOC00006
PS00008	12->18	MYRISTYL	PDOC00008
PS00008	14->20	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	47->53	MYRISTYL	PDOC00008
PS00008	166->172	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	218->224	MYRISTYL	PDOC00008
PS00008	222->228	MYRISTYL	PDOC00008
PS00008	234->240	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_20m24.2)

DKFZphut1_21d15

group: uterus derived

DKFZphut1_21d15 encodes a novel 191 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /chromosome="3"

Insert length: 5292 bp

Poly A stretch at pos. 5273, polyadenylation signal at pos. 5252

```
1 CTCCCACTAG TGTATGCCTT AATGGTGCCG CTCTTGTCGG CGTCTACGCT
51 TGGGACCTTG GCTTCTGACT TGGAGAGTGT ACAGCTCTGC CCGACGGCAA
101 CCCAGCTTGG GAAGAGAAGC CCCAGCGTGG GCTGGGGCTC AAGGCGCAGG
151 AAGGCCGAGC CCGGCGCGGA CGCAGGCGGC TCCGGGCGGG CTCAGCACCC
201 CCAGGCACCG TCTCCTAGTG ACCGCGGCGC TCGCGGGCCT GCGGGCCGTT
251 GTCCGGGCGA CTGCGCAGCG CGGGCACCCC GCGGGCCCTT CCCCTGGGCG
301 CGCGGCGGAC CTGGGTGCCA TGGCGGCAGC GCGGGTGACA GGCCAGCGGC
351 CTGAGACCGC GCGGCGCGAG GAGGCCTCGA GGCGGCAGTG GGCGCGGCCA
401 GACCACTGCC AGGCTCAGGC GCGGCGCGGG CTGGGCGAGC GCGAGGACGC
451 ACCGGTGCGT CCGCTGTGCA AGCCCCGCGG CATCTGCTCG CGCGCCTACT
501 TCCTGGTGCT GATGGTGTTT GTGCACCTGT ACCTGGGTAA CGTCTGGCG
551 CTGCTGTCTT TCGTGCACTA CAGCAACGGC GACGAAAGCA GCGATCCCCG
601 GCCCCAACAC CGTGCCCAAG GCCCGGGGCC CGAGCCCACT TTAGGTCCCC
651 TCACCCGGCT GGAGGGCATC AAGGTGAGGA CCTCCCTGCC CCGCCGCGCT
701 CCAGGCCCTG CACGGCTGAG CCCGAGAGGA CCGGCGCTCA GCGCGGGTCC
751 CCACGCTGCC CCGGCGCGTG CTCTGCGTCG GTCCCGCGCG CTCCCACTCA
801 CTCGCTGTCT GTCGCTCTCC GGGCGGGGCG GACTTGCCCC TTTTGGGCA
851 GCGCGGTCTG GCGCCCCAGC TGCCCGCTGT GCGCCTTTTC CTTAGGTGGG
901 GCACGAGCGT AAGGTCCAGC TGGTCACCGA CAGGGATCAC TTCATCCGAA
951 CCCTCAGCCT CAAGCCGCTG CTCTTCGAAA TCCCCGGCTT CCTGACTGAT
1001 GAAGAGTGTC GGCTCATCAT CCATCTGGCG CAGATGAAGG GGTTACAGCG
1051 CAGCCAGATC CTGCCTACTG AAGAGTATGA AGAGGCAATG AGCACTATGC
1101 AGGTGAGCCA GCTGGACCTC TTCCGGCTGC TGGACCAGAA CCGTGATGGG
1151 CACCTTCAGC TCCGTGAGGT TCTGGCCCAG ACTCGCCTGG GAAATGGATG
1201 GTGGATGACT CCAGAGAGCA TTCAGGAGAT GTACGCCGCG ATCAAGGCTG
1251 ATCCGTGATG TGACGGTGAG CTCACACCTC TGCACAGTCC TATCCCCGTG
1301 AGCCTCCTGC CCACTCCAGG GTGCACAATT TTGAAACTT GGGCCCTTCC
1351 CCCACAGCCA GGCAGCCTCT CTGCACCCCT TTATAGTGGC CAGAGATGGG
1401 GAGGTGAAGA TCCAGCCTTG CTTTTTACCC CTGGGAAGTA GGCAGGCAGC
1451 CAGGCCCCCC GTTCCCTTGG GTGATGGTCT CGAGGGCAGT TCTTGAGAGC
1501 GCTTTTGATA ACATCAGGCA GAGTTGAGAG CCTGGGGACA GGAAGTAGGG
1551 CTGCTAGTTG GCAGAGAACA GAGTGGGTGG AGCAGGAGCA AGGCGACAGT
1601 GAGGCCAGCT AGAGCTTGGC TGTTTACCC TCTCCATCCA TCTCTCCAGC
1651 CAGACACGAG GTCCACCCCA GCAGACAGCT TCCCTGGTCT AAGTGAGGTC
1701 TCCCTTGCC TCTCTTGTG CACCTGGAGT CATGCCGAAG CGCCTAAAT
1751 GGTAGTGCTG CTACCTGTGC TAACTGCTGG GGAGGGGTGG GCAGGGAAGC
1801 TGTCAATGCA GTGGTGCCCC CTCTGGTAAT AACTCTCAGG AGGTTTCTGA
1851 GGTGTGGTCA TCACCCTCAT GCCCAAATTC TGGACCAAGA GAGGAAGATA
1901 CAGCAGTTAG AAAGGACTTG GAACAGTGGC TTTGCGGCTG GTGAACCAGA
1951 GTGAAGAATC TGGCCGTGAC CTGGCTGCCA CACTGTCTATA GGCCCCAGAA
2001 CAGAGGTGGT GACAGTCTCA CAGCCCTTGA ATGTCCCCCA CCCTCAGAGG
2051 AATCTGGGCC AAAGAGTGGG AGGTGATGTC CTTGGGTCAG CCAGAATAAC
2101 ATGGAGCAAA GATACCAACT ACTCTTCCAG AACCCCAAGA GGGTAGAACC
2151 CCTGCTTAAT GGTTTGAGCA GGGACAGTGG AGAATGTTCT CATGAGAGGG
2201 GGTGGCCTGA CTTTCGTTGC TAAGTGGGCT GGTAACGCAG TAGGCAGGGC
2251 TGGCGAAGTA GGTTCACCCC AGGATGAAAC CTGGGGTCAT GAGGAACTCC
2301 CCGGGGGCTG GCCCTGCTTG CACCTGGCGG TATGTATGTA AGGCCCTGGA
2351 TGAGGCCCAG CACTGCCTGC TCTCTCTCA CCCTCCACAG GCCGGAGAGT
2401 GGCCACCACT CTATATAGCC AGGCTGGAAG GCCAGGGTCC TGGCCATATG
2451 GCTCAAGCTT CTTTGGGAGA ACCTTCTCTG GCCACTCTAA TAGGGGGTGG
2501 GCCTCTTTCT TCTTAGGGCC AAATTAGGGC TTAAACTGAG AAAAGGAACT
2551 GCTCTGGGTC TTCCTGTAAG GCCTGATGTG ACAGAAACCA GGTTCATCTG
2601 ACCCAAAAGT CCAGGTGGGG GACAAGTGTA CAAGGCCCTT CAGTGCCCTGA
2651 GGTACAGGGC TGCTGCTGCC TTTGGGGTAG GTAGGGGAAG GCAGCCTGCC
2701 ACTGTTGCC CTCAATATGG GCTTGGTGGG CATTGATGGT GGGTGCCCTG
2751 TGCAGGAGTG CTGAGTCTGC AGGAGTTCTC CAACATGGAC CTTCCGGGACT
2801 TCCACAAAGTA CATGAGGAGC CACAAGGCAG AGTCCAGTGA GCTGGTGGCG
```

```

2851 AACAGCCACC ATACCTGGCT CTACCAGGGT GAGGGTGCCC ACCACATCAT
2901 GCGTGCCATC CGCCAGAGGT GAGCACCTGA AGCTGTTCTC ACTGGAGCAG
2951 GGGGAGAAGA CTGGGCAGGG CCTCCACAGA AGTCCTTGTC TGGGGCCAAG
3001 AGGACAGAAT GGATTAACCC ATTTGGGATT AAGTTCCATT TGTTAGACCA
3051 GGATTGGGAC CCACGTAAAG ACAGGCAATT AACAAAGGCA AATTAGCCCT
3101 CCTTGCAGGC ACACAATGGG CAACTGGGGT TAGATAGAGA TTGAGCACTT
3151 CTTTCTGATT AGATAAATGA CCTCTTATCT TTGACCCCTT ATCTGACCCC
3201 GTCACAGCAG GAAAAGGGTT TTTAAATAAA CAACTTTCTT CCAGGGAGGA
3251 GGACCTCAGG ACTCCCCGCC CCCTTTATTT AGTGGAAATG TCAACATTTC
3301 CACATAGCAG GTGTCTCTGT CTTTGGCATC TGAGGGAGAA GGATCATCAT
3351 GAGTAACCCC CTCCTGCTCT TACAGGGCCA GTCTGAGATG GCTTAAGGGA
3401 CTTCCAGGGG AGGTGGGTAG GGGCAAAGCT TGTGGCAGGC CTAGGGTCCA
3451 CCTTGGCCAG CTCCTTCAGA TCACCACCTT GCCTGGGGCT GCCCAGCCAA
3501 ATGCTGCTG CCCACCAGGG TGCTGCGCCT CACTGCGCTG TCGCCTGAGA
3551 TCGTGGAGCT CAGCGAGCCG CTGCAAGTTG TTCGATATGG TGAGGGGGGG
3601 CACTACCATG CCCACGTGGA CAGTGGGCCT GTGTACCCAG AGACCATCTG
3651 CTCCCATACC AAGCTGGTAG CCAACGAGTC TGTACCCTTC GAGACCTCCT
3701 GCCGGCAAGT ATCTCCCAAC TGGGGGCTGC CTTCAATCCT CAGACCAGGA
3751 ACACCCATGA CACAGGCACA GCCCTGCACT GTGGGCGTGC CCCTTGGCAT
3801 GGGGGCAGGA GATCACTGGG TTATCCCGGT TAGTGATGCC CTCACCTCTC
3851 CCCACAAGTT GTTTACCCAA TGGCTGGAAA GGGGTGGCTA CTGGTCATCG
3901 TGACCACTGG AGTCAACACA GACTGATGTA CCCACAGACA CCAAAACTTG
3951 CCCCCGTAGT TCTGAAGCAA GGGGCAAGGC TGGGGCCCTA GCTTGTCTCT
4001 CCCATTCCTC CAGGTGTTGA TCTTGATTCC ACTTAGAGAA GCTGAAGCTG
4051 TGCCTCCCTC CCCTGTCAAG CCAGTTCTTT CCTCTTCAGG TGGCTGTTCT
4101 GGCCGAGCCC CTTCCCATCC CCAAGGAGCC CTTACGCGCG CCCTGTTGCT
4151 TCTGCTAGCC TACCTTTCCC TGCCAGGCCG TTGCTCAGGG CCATGGCATT
4201 TAACTAAGTG CACCTGTGAT CTTGGCCAAA AAACCATTGC AACTCACAGT
4251 AAGAGACTGG GTTTCGGGGA AGGAGGGGCT AGGGACATTT TGGCACTGGC
4301 CTGCCCTATT GTCTCCCATC CTAGTCTGTC CTGGTCCCTG GCAACAGGAA
4351 CCTGGGCAGC TTATCCTGCC CACAGGTAAAG CCCCTGGGAG CATCCACAAC
4401 TGGGGACCTG CTCAGTGCCC CCCCTGCCTT ACAGCTACAT GACAGTGCTG
4451 TTTTATTGTA ACAACGTAC TGGTGGGGGC GAGACTGTTT TCCCTGTAGC
4501 AGATAACAGA ACCTACGATG AAATGGTAAG GGTCAACTGG GCTATTACTC
4551 TTGTGGGCTG GCAGGGGCTT AGACAAGTGA AGTACACACC TCTCCAGGTC
4601 TAAGGATGTG GGCCCAAATT ATTCTTTGGG CATATCTGGT TGGTTTCCCT
4651 TTGCTACCCC TTGGCTGGCC TGGCCATAGA GTGGGGACAG GTTGAACACC
4701 CCACCACCTT GCTGCCACA GAGTCTGATT CAGGATGACG TGGACCTCCG
4751 TGACACACGG AGGCACTGTG ACAAGGGAAC CCTGCGTGTC AAGCCCCAAC
4801 AGGGCACAGC AGTCTTCTGG TACAACTACC TGCCTGATGG GCAAGGTTGG
4851 GTGGGTGACG TAGACGACTA CTCGCTGCAC GGGGGCTGCC TGGTCACGCG
4901 CGGCACCAAG TGGATTGCCA ACAACTGGAT TAATGTGGAC CCCAGCCGAG
4951 CGCGGCAAGC GCTGTTCCTA CAGGAGATGG CCCGCTTGC CCGAGAAGGG
5001 GGCACCCGACT CACAGCCCGA GTGGGCTCTG GACCGGGCCT ACCGCGATGC
5051 GCGCGTGGA CTCTGAGGGA AGAGTTAGCC CCGGTTCCTA GCCGCGGGTC
5101 GCCAGTTGCC CAAGATCAGG GGTCCGGCTG TCCTTCTGTC CTGCTGCAGA
5151 CTAAAGGTCT GGCCAATGTC TTGCCCAACC CCGCCAGCCG CGATACGGCG
5201 CAGTTCCTAT ATTATGTTA TTTATTGTGT ACTGACTCCA TCTGCCCCGT
5251 CAAATAAAAA ACCACAAGGT TCGAAAAAAA AAAAAAAGG GG

```

BLAST Results

Entry HSU64252 from database EMBL:
 Human STS sequence NOTI-225.
 Score = 959, P = 1.2e-36, identities = 195/199

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 351 bp; peptide length: 118
 Category: questionable ORF
 Classification: no clue

```

1 LPLVYALMVP LLSASTLGTL ASDLESVQLC PTATQLGKRS PSVWGSRRR
51 KAEPGADAGG SGRAQHPQAP SPSDRGARGP GGRCPGDCAA RAPPRPLPWA
101 RARPGCHGGS GGD RPAA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 1

No Alert BLASTP hits found

Peptide information for frame 2

ORF from 320 bp to 892 bp; peptide length: 191
Category: putative protein
Classification: no clue

```

1 MAAAVTGQR PETAAEEAS RPQWAPPDHC QAQAAAGLGD GEDAPVRPLC
51 KPRGICSRAY FLVLMVFVHL YLGNVLALLL FVHYSNGDES SDPGPQHRAQ
101 GPGPEPTLGP LTRLEGIKVR TSLPRRAPGP ARLSPRGPAL SPGPHAAPGA
151 ALRRSRALPL TRLLSLSGPG RLGPFWAARS GAPAARCAPF P

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_21d15, frame 2

PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1, N = 2,
Score = 106, P = 0.0067

>PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1
Length = 1,298

HSPs:

Score = 106 (15.9 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 36/103 (34%), Positives = 44/103 (42%)

```

Query: 87 GDESSDPGQHRAQGGPPEPTLGLTRLEGIKVRTSLPRRA-PGPARLS-PRGPALSPGP 144
      G + PGP G GP P P T+ G S R P PA S P GP +P
Sbjct: 726 GKRRKSPGPARPPGGGGPRP---PKTKKSGADAPGSDARAPLPAPAPPSTPPGPEPAPAQ 782

```

```

Query: 145 HAAPGAALRRSRALPLT-RLLSLSGPGRLGPFWAARSGAPAARCAP 189
      AAP AA ++R P+ GP LG W + P+ AP
Sbjct: 783 PAAPRAAAQARPRPVAVSRPFAEGPDPLGG-WRRQPPGPSHTAAP 827

```

Score = 40 (6.0 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
Identities = 8/21 (38%), Positives = 9/21 (42%)

```

Query: 28 DHCQAQAAAGLGDGEDAPVRP 48
      DH + A G G AP P
Sbjct: 212 DHAREARAVGRGPSSAAPAAP 232

```

Pedant information for DKFZphut1_21d15, frame 1

Report for DKFZphut1_21d15.1

```

[LENGTH] 117
[MW] 11797.32
[pI] 10.68
[KW] Irregular
[KW] SIGNAL_PEPTIDE 22
[KW] LOW_COMPLEXITY 38.46 %

```

```

SEQ LPLVYALMVLLSASTLGTLASDLESVQLCPTATQLGKRSPSVGWGSRRRKAEPGADAGG
SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD cccccccccccccccccccccchhhhhhhcccccccccccccccccccccccccccccccc

```

```

SEQ SGRAQHPQAPSPSDRGARGPGGRCPGDCAARAPPRLPWARARPGCHGSGGSDRPAA
SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

(No Prosite data available for DKFZphut1_21d15.1)

(No Pfam data available for DKFZphute1_21d15.1)

Pedant information for DKFZphute1_21d15, frame 2

Report for DKFZphute1_21d15.2

{LENGTH} 191
{MW} 19916.88
{pI} 10.43
{KW} TRANSMEMBRANE 1
{KW} LOW_COMPLEXITY 29.84 %

SEQ MAAAVTGQRPETAAEEASRPQWAPPDHCQAQAAAGLGDGEDAPVRPLCKPRGICSRAY
SEG
PRD cccceeeccccchhhhhhhhhccccccchhhhhhhccccccccccccccccccccchhhh
MEM

SEQ FLVLMVFVHLYLGNVLALLLFVHYSNGDESSDPGPQHRAQGPPEPTLGPLTRLEGIKVR
SEGxxxxxxxxxxxxxxxxxxxx.....
PRD hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccceeeee
MEMMMMMMMMMMMMMMM.....

SEQ TSLPRRAPGPAPRLSPRGPALSPGPHAAPGAALRRSRALPLTRLLSLSGPGRLGPFWAARS
SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD eccccccccccccccccccccccccccccchhhhhhhccccceccccccccchhhhhhc
MEM

SEQ GAPAAARCAPPF
SEG xxxxxxxxxx..
PRD ccccccccccc
MEM

(No Prosite data available for DKFZphute1_21d15.2)

(No Pfam data available for DKFZphute1_21d15.2)

DKFZphute1_22d2

group: signal transduction

DKFZphute1_22d2 encodes a novel 580 amino acid putative GTP-binding protein related to the ras protein. Additionally, the putative protein contains an EF-hand for calcium-binding.

G-proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to GTP-binding proteins

complete cDNA, complete cds, potential start at Bp 64, EST hits
complete cds according to K08F11.5 and YAL048c

Sequenced by BMFZ

Locus: /map="17"

Insert length: 3247 bp

Poly A stretch at pos. 3230, no polyadenylation signal found

```
1 CTCCTGGTGA GAGGAGTCCA CTCCTGCGT GCGGGCGGAG GCCGGCCCCC
51 GAGAGCCGCC GACATGAAGA AAGACGTGCG GATCCTGCTG GTGGGAGAAC
101 CTAGAGTTGG GAAGACATCA CTGATTATGT CTCCTGGTCAG TGAAGAATTT
151 CCAGAAAGAGG TTCCTCCCCG GGCAGAAAGAA ATCACCATTG CAGCTGATGT
201 CACCCAGAG AGAGTTCCAA CACACATTGT AGATTACTCA GAAGCAGAAC
251 AGAGTGATGA ACAACTTCAT CAAGAAATAT CTCAGGCTAA TGTCTCTGT
301 ATAGTGTATG CCGTTAACAA CAAGCATTCT ATTGATAAGG TAACAAGTCG
351 ATGGATTCTT CTCATAAATG AAAGAACAGA CAAAGACAGC AGGCTGCCTT
401 TAATATTGGT TGGGAACAAA TCTGATCTGG TGGAAATAG TAGTATGGAG
451 ACCATCCTTC CTATTATGAA CCAGTATACA GAAATAGAAA CCTGTGTGGA
501 GTGTTACGCG AAAAACCTGA AGAACATATC AGAGCTCTTT TATTACGCAC
551 AGAAAGCTGT TCTTCATCCT ACAGGGCCCC TGTACTGCCC AGAGGAGAAG
601 GAGATGA AAC CAGCTTGTAT AAAAGCCCTT ACTCGTATAT TTAATATATC
651 TGATCAAGAT AATGATGGTA CTCCTCAATGA TGCTGAACCT AACTTCTTTC
701 AGAGGATTTG TTTCAACACT CCATTAGCTC CTCAGCTCTT GGAGGATGTC
751 AAGAATGTAG TCAGAAAACA TATAAGTGAT GGTGTGGCTG ACAGTGGGTT
801 GACCTGAAA GGTTTTCTCT TTTTACACAC ACTTTTATAT CAGAGAGGGA
851 GACACGAAAC TACTTGGACT GTGCTTCGAC GATTGGGTTA TGATGATGAC
901 CTGGATTGTA CACCTGAATA TTTGTTCCCT CTGCTGAAAA TACCTCCTGA
951 TTGCACTACT GAATTAATC ATCATGCATA TTTATTCTC CAAAGCACCT
1001 TTGACAAGCA TGATTGGAT AGAGACTGTG CTTTGTCACT TGATGAGCTT
1051 AAAGATTAT TTAAGTTT CCCTTACATA CCTTGGGGGC CAGATGTGAA
1101 TAACACAGTT TGTACCAATG AAAGAGGCTG GATAACCTAC CAGGGATTCC
1151 TTTCCAGTG GACGCTCAGC ACTTATTAG ATGTACAGCG GTGCTGGGAA
1201 TATTTGGGCT ATCTAGGCTA TTCAATATTG ACTGAGCAAG AGTCTCAAGC
1251 TTAGCTGTT ACAGTGACAA GAGATAAAAA GATAGACCTG CAGAAAAAAC
1301 AAACCTCAAG AAATGTGTTC AGATGTAATG TAATTGGAGT GAAAAACTGT
1351 GGGAAAGTGG GAGTTCTTCA GGCTCTTCTT GGAAGAAACT TAATGAGGCA
1401 GAAGAAAATT CGTGAAGATC ATAAATCCTA CTATGCGATT AACACTGTTT
1451 ATGTATATGG ACAAGAGAAA TACTTGTGTG TGCATGATAT CTCAGAATCG
1501 GAATTTCTAA CTGAAGCTGA AATCATTTGT GATGTTGTAT GCCTGGTATA
1551 TGATGTCAGC AATCCCAAAT CCTTGAATA CTGTGCCAGG ATTTTAAAGC
1601 AACACTTTAT GGACAGCAGA ATACCTTGCT TAATCGTAGC TGCAAAGTCA
1651 GACCTGCATG AAGTTAAACA AGAATACAGT ATTTACACCTA CTGATTTCTG
1701 CAGGAAACAC AAAATGCCTC CACCACAAGC CTTCACTTGC AATACTGCTG
1751 ATGCCCCAG TAAGGATATC TTTGTTAAAT TGACAACAAT GGCCATGTAT
1801 CCGTAAGTAC TTGCTGTCTT CATTTTCATG TTGCTAGGTT CATAACATTG
1851 CATGCCATTA TTAGCCATGA AGGGAATATC TTTGTACAT AGGAATTGTT
1901 CAGCAACAGA AAGATACTTT GTAATGAGAA GGTACAAATT TGAGTAAATG
1951 CAAGTTTGGT TTGAATGCCA TAATAAAATG ATATAACAG TGCTTCTGAC
2001 AATATCTGTA TATTTTGGAG CAGGCTGTAA CTATCTTAAT AGAATAGTAC
2051 AAATAAACAC AACCCCCAC CCAGCATTA AAAATAGTTT TACTGGAATA
2101 AATGGGTTT GGCATCATGT TGTTTTATGC TTATAAAGCA TTTTCATATG
2151 AACAGAAAGT TTATATTTT CTGTTTTTGA CCTTAGGTAT ATGAAGTTT
2201 CTAAAAATAT TTATTAATTT ATGTTGAAAT TGTGGGTATG CTTAGTTAG
2251 GATATGCTT TTTTAAAGTG TGTAAGAGT AGTTGTAATT GGAATTTCTA
2301 CTGTATAAAT GTTTTACATT AAGTGTTACG AGCCACAAT TTCATGTACA
2351 TTTATTATAT ATCTATACAT GCATATGCAC AAGCACATAA CTGTGGTCAT
2401 CTCGTGATGT TACTAACTGC CTTAAAAATG CATGGTTCTT AATGGCATTC
2451 GCCTCAAGTA GTGTGTTTGT ATAAATCTG TTTTGTAAAC AAATAGTTT
2501 TCAGGCAGTG CGTTTCTCAG GACTTTATAG CTTATTCTAC TTATTCTTAT
2551 GTTAGTCTCT AAATTATTTT TCTTCTTATG AAAACTACAG TGAACACAG
```

```

2601 AGTAATAATC AAACATTGCT ATAAACCAAG AATGACATTT TTCAAAAAGG
2651 TGTTGATTGG TACAGATTTT TAAAGTCAGT TAACTTTACT GCTATTTTAT
2701 TACCTAATAC TTTTTTTAGA TGCAACAAAC CCTTGAATTT CTATTGTAT
2751 TCGAAGACAA GTCATTCCCTA TTATTATAGA ATAACCAAAA CCTTATTAT
2801 GTTTTACCTT TGCTTTAAAA CTCTCATGTA TGTATCTAC AGAGAGGATC
2851 ATTACAGAGA CAGACTCTCC CGAGACATGG GCCACACTGA TAGAATAGAG
2901 AATTTGAGAA AAATCTGGGT CTTTCTAAAA ACTGCTTTGT AAGTTACTTT
2951 TTCTTTATGA CTTCTGTGGG ATTTTGTGTA TATTTTCTTA GAGAATGACC
3001 AAATCTCCTT TCTTGCCATA ATTAACATTT AGTAATTATG TAGAAACGCA
3051 CTGCTTGGTC AGGCTTCCTG CCTAGCTATA TATTACGTTG TCTTCCTTAC
3101 TACATAAATG TACTTCTTTA ATCTTGTGAT TACAGTAACT GCAAGTGTGT
3151 TTTTACATCT GCATTTTAA AACATTTTAC TGTAAATTCTG TTGTGTGTGT
3201 GTGTGTTATA TGATAAATGT ACATACATGG AAAAAAAAAA AAAAAAA

```

BLAST Results

Entry AC004527 from database EMBL:
 *** SEQUENCING IN PROGRESS *** NF1-related locus, Direct Submission;
 HTGS phase 1, 10 unordered pieces.
 Score = 1899, P = 1.1e-78, identities = 387/396

Entry HS148355 from database EMBL:
 human STS SHGC-31220.
 Score = 1826, P = 7.5e-78, identities = 388/406

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 64 bp to 1803 bp; peptide length: 580
 Category: similarity to known protein

```

1 MKKDVRILLV GEPRVGKTSL IMSLVSEEF EEVPPRAEEI TIPADVTPER
51 VPTHIVDYSE AEQSDQLHQ EISQANVICI VYAVNNKHSI DKVTSRWIPL
101 INERTDKDSR LPLILVGNKS DLVEYSSMET ILPINQYTE IETCVCESAK
151 NLKNISELFY YAKAVLHPT GPLYCPEEKE MKPACIKALT RIFKISDQDN
201 DGTLNDAELN FFQRCFNTF LAPQALVDK NVVRKHISDG VADSGTLKLG
251 FLFLHTLFIQ RGRHETTWTV LRRFGYDDDL DLTPEYLFPL LKIPPDCTTE
301 LNHAYLFLQ STFDKHDLDL DCALSPDELK DLKFVFPYIP WGPDVNNTVC
351 TNERGWITYQ GFLSQWTLTT YLDVQRCLEY LGYLGYSILT EQESQASVT
401 VTRDKIDLQ KKQTQRNVFR CNVIGVKNCG KSGVLQALLG RNLMRQKKIR
451 EDHKSYYAIN TVVYVQGEKY LLLHDISESE FLTEAEIICD VVCLVYDVSN
501 PKSFEYCARI FKQHFMSRI PCLIVAAKSD LHEVKQEYSI SPTDFCRKHK
551 MPPPQAFTCN TADAPSKDIF VKLTTMAMYP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutell_22d2, frame 1

TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11., N = 1, Score = 1357, P = 1.1e-138

TREMBL:SPCC320_4 gene: "SPCC320.04c"; product: "hypothetical protein";
 S.pombe chromosome III cosmid c320., N = 1, Score = 889, P = 4.4e-89

TREMBL:CEUC47C12_3 gene: "C47C12.4"; Caenorhabditis elegans cosmid
 C47C12., N = 2, Score = 408, P = 5.6e-74

PIR:S51971 probable membrane protein YAL048c - yeast (Saccharomyces
 cerevisiae), N = 1, Score = 677, P = 1.3e-66

>TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
 K08F11.

Length = 625

HSPs:

Score = 1357 (203.6 bits), Expect = 1.1e-138, P = 1.1e-138
Identities = 263/582 (45%), Positives = 380/582 (65%)

Query: 4 DVRILLVGEPRVGKTSLIMSLVSEEFPEEVPPRAEEITIPADVTPEVPTHIVDYSEAEQ 63
DURI+L+G+ GKTSL+MSL+ +E+ + VP R + + IPADVTPE V T IVD S E+
Sbjct: 9 DVRIVLIGDEGCGKTSLVMSLLEDEWDAVPRRLDRVLIPADVTPEVVTTSIVOLSIKEE 68

Query: 64 SDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKSDLV 123
+ + EI QANVIC+VY+V ++ ++D + ++W+PLI + + P+ILVGNKSD
Sbjct: 69 DENWIVSEIRQANVICVVSVTDESTVDGIQTKWLPLIRQSFGEYHETPVILVGNKSDGT 128

Query: 124 EYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYYAQKAVLHPTGPLYCPEEKEMKP 183
++ + ILPIM TE+ETCVECSA+ +KN+SE+FYAQKAV++PT PLY + K++
Sbjct: 129 A-NNTDKILPIMEANTEVETCVECSARTMKNVSEIFYAQKAVIYPTRPDYADTKQLTD 187

Query: 184 ACIKALTRIFKISDQDNDGTLDNAELNFFQRICFNTPLAPQALEDVKNVVRKHISDGVAD 243
KAL R+FKI D+DNDG L+D ELN FQ++CF PL ALEDVK V DGVA+
Sbjct: 188 RARKALIRVFKICDRDNDGYLSDTELNDFFQKLCFGIPLTSTALEDVKRAVSDGCPDGVAN 247

Query: 244 SGLTLKGFLEHTLFIQGRHETTWTVLRFRGYDDDLDTPEYLFPLLKIPDCTTELNH 303
L L GFL+LH LFI+RGRHETT WLR+FGY+ L L + +YL+P + IP C+TEL+
Sbjct: 248 DSLMLAGFLYLHLLFIERGRHETTWAVALRKFYETSLKLSSELYPRITIPVGCSTELSP 307

Query: 304 HAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVNNVTCTNERGWITYQGFL 363
F+ + F+K+D D+D LSP EL++LF V P D + TN+RGW+TY G++
Sbjct: 308 EGVQFVSALFEKYDEDKDGLSPSELQNLFSVCPVPVITKDNILALETNQRGWLTYNGYM 367

Query: 364 SQWTLTTYLDVQRCLEYLGLYLSILTEQESQAS----AVTVTRDKKIDLQKKQTRNVF 419
+ W +TT +++ + E L YLG+ + +A ++ VTR++K DL+ T R VF
Sbjct: 368 AYWNMTTLINLTQTFEQLAYLGFVPVGRSGPGRAGNTLDSIRVTRERKKDLENHGTDRKVF 427

Query: 420 RCNVIGVKNCGKSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYQKEYLLHLDI--- 476
+C V+G K+ GK+ +Q+L GR + +I H S + IN V V + KYLLL ++
Sbjct: 428 QCLVVGAKDAGKTVMQSLAGRGMDVAQIGRRH-SPPFVINRVRVKEESKYLLREVDVL 486

Query: 477 SESEFLTEAEIICDVVCLVYDVSNPKSEFYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQ 536
S + L E DVV +YD+SNP SF +CA +++++F ++ PC+++A K + EV Q
Sbjct: 487 SPQDALGSGETSADVVAFLYDISNPDSFAFCATVYQKYFYRTKTPCVMIATKVEREEVDQ 546

Query: 537 EYSISPTDFCRKHKMPPPOAFTCNTADAPSKDIFVKLTMMAMP 580
+ + P +FCR+ ++P P F+ S IF +L MA+YP
Sbjct: 547 RWEVPPPEEFCRQFELPKPIKFSTGNIGQSSSPIFEQLAMMAVYP 590

Pedant information for DKFZphutel_22d2, frame 1

Report for DKFZphutel_22d2.1

[LENGTH] 580
[MW] 66541.61
[pI] 5.56
[HOMOL] TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11. 1e-149
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL048c] 5e-81
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR055w] 3e-11
[FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 8e-09
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 4e-08
[FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 4e-08
[FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YPR165w] 7e-08
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 9e-08
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YNL093w] 1e-07

[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YNL093w]
 1e-07
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YNL093w] 1e-07
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-07
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
 [S. cerevisiae, YLR229c] 8e-07
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 3e-06
 [FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
 YGL210w] 9e-04
 [BLOCKS] BL00410A Dynamin family proteins
 [SCOP] dlplk_ 3.25.1.3.1 CH-p21 Ras protein [human (Homo sapiens)] 2e-42
 [SCOP] dlguaa_ 3.25.1.3.10 Rap1A [Human (Homo sapiens)] 5e-59
 [PIRKW] transmembrane protein 1e-79
 [PIRKW] membrane trafficking 2e-06
 [PIRKW] acetylated amino end 3e-09
 [PIRKW] prenylated cysteine 3e-09
 [PIRKW] signal transduction 1e-07
 [PIRKW] transforming protein 3e-09
 [PIRKW] immediate-early protein 8e-06
 [PIRKW] alternative splicing 4e-08
 [PIRKW] P-loop 1e-10
 [PIRKW] lipoprotein 7e-10
 [PIRKW] proto-oncogene 3e-09
 [PIRKW] methylated carboxyl end 3e-09
 [PIRKW] membrane protein 3e-09
 [PIRKW] GTP binding 1e-10
 [PIRKW] thiolester bond 7e-10
 [SUPFAM] ras transforming protein 1e-10
 [PROSITE] ATP_GTP_A 2
 [PROSITE] MYRISTYL 3
 [PROSITE] EF_HAND 1
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 14
 [PROSITE] TYR_PHOSPHO_SITE 4
 [PROSITE] PKC_PHOSPHO_SITE 5
 [PROSITE] ASN_GLYCOSYLATION 3
 [PFAM] Ras family (contains ATP/GTP binding P-loop)
 [KW] Irregular
 [KW] 3D

SEQ MKKDVRILLVGEPRVGKTSLIMSLVSEEFPEEVPRAEEITIPADVTPERVPTHIVDYSE
 1jai- ...EEEEEEETTTTCHHHHHHHHHCCCCCCCCCEEEEEETEEEEEEEEEECCC

SEQ AEQSDQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKS
 1jai- CGGGHHHHHHHHHTTEEEEEETTTTHHHHHHH-HHHHHHHHHHCTTT-TCEEEEEETT

SEQ DLVEYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYQAQAVLHPTGPLYCPPEKE
 1jai- TTTTTTTTHHHHHHHHHHCCCE-EECTTTTTTHHHHHH.....

SEQ MKPACIKALTRIFKISDQNDGTLNDAELNFFQRICFNTPLAPQALDVKNVVRKHISDG
 1jai-

SEQ VADSGTLKGLFLHTLFIQGRHETTWTVLRFGYDDDLDTPEYLFPLLKIPPDCTTE
 1jai-

SEQ LNHHAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVNNVTCTNERGWITYQ
 1jai-

SEQ GFLSQWTLTTYLDVQRCLEYLGYLGYSILTEQESQASAVTVTRDKKIDLQKKQTQRNVFR
 1jai-

SEQ CNVIGVKNCGSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLHDISESE
 1jai-

SEQ FLTEAEIICDVVCLVYDVSNPKSFEYCARIKQHFMDSRIPCLIVAASDLHEVKQEYSI
 1jai-

SEQ SPTDFCRKHKMPPQAFTCNTADAPSKDIFVKLTMMAMP
 1jai-

Prosite for DKFZphut1_22d2.1

PS00001	118->122	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00004	411->415	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005

PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	311->315	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00007	153->161	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->384	TYR_PHOSPHO_SITE	PDOC00007
PS00007	153->162	TYR_PHOSPHO_SITE	PDOC00007
PS00007	448->457	TYR_PHOSPHO_SITE	PDOC00007
PS00008	240->246	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008
PS00008	433->439	MYRISTYL	PDOC00008
PS00017	11->19	ATP_GTP_A	PDOC00017
PS00017	425->433	ATP_GTP_A	PDOC00017
PS00018	197->210	EF_HAND	PDOC00018

Pfam for DKFZphut1_22d2.1

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK		
Query	6	RILLVGEPRVGKTSLIMSLVSEEFPEE-VPPR-AEEITIPADVTPERVP	52
HMM	LQIWDTAGQERYRsmRPMYYRGAMGFMLVYDITNRqSFENIr.NWweEIr		
Query	53	THIVDYSEAEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLIN	102
HMM	RHCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKT		
Query	103	ERTDKDSRLPLILVGNKSOLVEYSSMETILPIMNQYTEI-ETCVECSAKN	151
HMM	NiNVEEAFMEIvReIlqrMqeqNqteNinidQpsrnrkrCCCIM*		
Query	152	LKNISELFYYAQAVLHPT-----GPLYCPEEKEMK-PACI--	186

DKFZphut1_22e12

group: signal transduction

DKFZphut1_22e12 encodes a novel 92 amino acid protein, with similarity to yeast, C.elegans, Drosophila and mammalian proteins.

The Drosophila cni and mammalian cornichon proteins are part of a signal transduction pathway involving the EGF-receptor.

The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

strong similarity to S.cerevisiae YGL054c and cornichon

complete cDNA, complete cds, EST hits
cornichon is required for signal transduction in the EGF-receptor
signal processing

Sequenced by BMFZ

Locus: unknown

Insert length: 519 bp
Poly A stretch at pos. 499, no polyadenylation signal found

```

1  GTCGGGGCAT CCGAGCGGGT TTGACGGAAG GAGCGGCGGC GACGGAGGAG
51 GAGGATGGAG GCGGTGGTGT TCGTCTTCTC TCTCCTCGAT TGTTGCGCGC
101 TCATCTTCCT CTCGGTCTAC TTCATAATTA CATTGTCTGA TTTAGAATGT
151 GATTACATTA ATGCTAGATC ATGTTGCTCA AAATTAAACA AGTGGGTAAT
201 TCCAGAAATG ATTGGCCATA CCATTGTCAC TGTATTACTG CTCATGTCAT
251 TGCACTGGTT CATCTTCCTT CTCAACTTAC CTGTTGCCAC TTGGAATATA
301 TATCGTATGA TCTTAGCTTT GATAAATGAC TGAAGCTGGA GAAGCCGTGG
351 TTGAAGTCAG CCTACACTAC AGTGCACAGT TGAGGAGCCA GAGACTTCTT
401 AAATCATCCT TAGAACCCTG ACCATAGCAG TATATATTTT CCTCTTGGA
451 CAAAAAATA TTTTGCTGT ATTTTACCA TATAAAGTAT TAAAAAACA
501 TGAACAAAAA AAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

95300228:
cornichon and the EGF receptor signaling process are necessary for both
anterior-posterior
and dorsal-ventral pattern formation in Drosophila.

Peptide information for frame 1

ORF from 55 bp to 330 bp; peptide length: 92
Category: strong similarity to known protein

```

1 MEAVVFVFSL LDCCALIFLS VYFIITLSDL ECDYINARSC CSKLNKWWIP
51 ELIGHTIVTV LLLMSLHWFI FLLNLPVATW NIYRMILALI ND

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22e12, frame 1

PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae), N = 2, Score = 185, P = 5.7e-17

TREMBL:SPAC2C4_5 gene: "SPAC2C4.05"; product: "cornichon homolog";

S.pombe chromosome I cosmid c2C4., N = 1, Score = 163, P = 3.7e-12

PIR:S46084 probable membrane protein YBR210w - yeast (Saccharomyces cerevisiae), N = 1, Score = 162, P = 4.8e-12

TREMBL:AF104398_1 product: "cornichon"; Homo sapiens cornichon mRNA, complete cds., N = 1, Score = 141, P = 8e-10

SWISSPROT:CNI_DROVI CORNICHON PROTEIN., N = 1, Score = 139, P = 1.3e-09

>PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)
Length = 138

HSPs:

Score = 185 (27.8 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 35/85 (41%), Positives = 56/85 (65%)

Query: 1 MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWWIPELIGHTIVTV 60
M A +F+ +++ C +F V+F I +DLE DYIN CSK+NK + PE H +++
Sbjct: 1 MGAWLFILAVVNCINLFGQVHFTILYADLEADYINPIELCSKVNKLITPEALHGALSL 60

Query: 61 LLLMSLHWFIPLLNPVATWNIYRM 85
L L++ +WF+FLNLPV +N+ ++
Sbjct: 61 LFLNGYWFVFLNLPVLAYNLNKI 85

Score = 37 (5.6 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
Identities = 7/9 (77%), Positives = 9/9 (100%)

Query: 82 IYRMILALI 90
+YRMI+ALI
Sbjct: 123 LYRMIMALI 131

Pedant information for DKFZphut1_22e12, frame 1

Report for DKFZphut1_22e12.1

[LENGTH] 92
[MW] 10614.98
[pI] 5.04
[HOMOL] PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)
5e-14
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGL054c]
2e-15
[PIRKW] transmembrane protein 2e-11
[PROSITE] CK2_PHOSPHO_SITE 3
[KW] SIGNAL_PEPTIDE 33
[KW] TRANSMEMBRANE 2

SEQ MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWWIPELIGHTIVTV
PRD ccchhhhhhhhhhhhhhhhhheeeccccccccccccccccceehhhhhhhhhhh
MEMMMMMMMMMMM

SEQ LLLMSLHWFIPLLNPVATWNIYRMILALIND
PRD hhhhhhhheeeccccchhhhhhhhhhhhhccc
MEM MMMMMMMMMMMMMMMMM..MMMMMM...

Prosite for DKFZphut1_22e12.1

PS00006 9->13 CK2_PHOSPHO_SITE PDOC00006
PS00006 26->30 CK2_PHOSPHO_SITE PDOC00006
PS00006 28->32 CK2_PHOSPHO_SITE PDOC00006

(No Pfam data available for DKFZphut1_22e12.1)

DKFZphutel_22n2

group: uterus derived

DKFZphutel_22n2 encodes a novel 304 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="553.3 cR from top of Chr11 linkage group"

Insert length: 1556 bp

Poly A stretch at pos. 1534, no polyadenylation signal found

```
1  ACAACAGGCT GGTTCCTTGG CGTGGAAATCC TAAAGTGGCC TGGCTTTGAG
51  ACTGGAGTGA GACCCCAGCC CTAGGCTGGG GTTCTTTCCA TTATAGAGGA
101 GACGGATTCA GAAGGGCTAC AGACCAAGGT TGTGAAAAC CAGACATATG
151 ATGAGCGTCT AGAGATTAAAC GACTCCGAAG AGGTTGCAAG TATTTATACT
201 CCAACCCCAA GACACCAAGG ACTTCCTCGT TCTGCCATC TTCCTAACAA
251 GGCTATGGCT GATAACAGCA GTGATGAGTG TGAAGAGGAA AATAACAAGG
301 AGAAGAAGAA GACCTCACAG TTGACACCTC AACGGGGCTT TAGTGAAAAT
351 GAGGATGACG ATGATGATGA TGATGATTCA TCTGAAACTG ATTCTGATTG
401 TGATGATGAT GATGAAGAGC ATGGAGCCCC TCTGGAAGGG GCCTATGACC
451 CTGCAGACTA TGAGCATTTG CCAGTTTCTG CTGAAATTAA GGAACCTCTC
501 CAGTACATCA GTAGGTACAC ACCTCAGTTG ATTGACCTGG ACCACAAACT
551 GAAGCCTTTC ATTCTTGATT TTATCCCAGC TGTCGGGGAT ATTGATGCAT
601 TCTTTAAAGT CCCACGTCCT GATGGAAAAGC CTGACAACCT TGGCCTATTG
651 GTATTGGATG AACCTTCTAC AAAGCAGTCA GACCCCTACG TGCTCTCACT
701 CTGGTTAAAC GAGAATTCTA AGCAGCACAA CATCACACAA CATATGAAAG
751 TAAAAAGCCT AGAAGATGCA GAAAAGAATC CCAAAGCCAT TGACACGTGG
801 ATTGAGAGCA TCTCTGAATT ACACCGTTCT AAGCCCCCTG CGACTGTGCA
851 CTACACCAGG CCCATGCCCG ACATTGACAC GCTGATGCAG GAATGGTCCC
901 CGGAGTTTGA AGAGCTTTTG GGCAAGGTAA GCCTGCCCAC GGCAGAGATT
951 GATTGCAGCC TGGCAGAGTA CATTGACATG ATCTGTGCCA TTCTAGACAT
1001 CCCTGTCTAC AAGAGTCGGA TCCAGTCCCT CCATCTGCTC TTTTCCCTCT
1051 ACTCAGAAAT CAAGAACTCA CAGCATTTTA AAGCTCTCGC TGAAGGCAAG
1101 AAAGCATTCA CTCCTTCATC CAATTCCACC TCCCAAGCTG GAGACATGGA
1151 GACATTAAAC TTCAGCTGAG ACACTTCCCA AGCTGCTGTT TCAAGGCTGA
1201 GCTGGCCCTC CTGCCCCAGC TGAGATGGAC AGATCGTTGT CAGCTACTTG
1251 ATGTCCTTGC CCATGCCACA GCTTGCTCA GGGGCAGTGC ATGTCCTGCT
1301 GCCCTCTCTG CCAGAGGGCA CAGAACATGT TTGTTTAATG AACCTGCCTG
1351 CCTCAGATTG CTGTCCCCGG GGAGTTAATG CATCTACACC ACTGTGGGGA
1401 TTTGAGTTAT AAGAAATTGA ATTTCTGAGA TCCCATGGAG GTTAGATTGG
1451 GAGGAAAGCT TAAAAGATGT CCTTTTGTG AGAGGGATGG AATTGTTTTT
1501 TTTCATTCTG AAAGTTAGTG AGTAAAGATT TTATAAATCA AAAAAAAAAA
1551 AAAAAA
```

BLAST Results

Entry HS188252 from database EMBL:

human STS WI-12265.

Score = 2554, P = 4.1e-109, identities = 556/587

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 255 bp to 1166 bp; peptide length: 304

Category: putative protein

```

1  MADNSSDECE EENNKEKKKT SQLTPQRGFS ENEDDDDDDD DSSETDSDS
51  DDDEEHGAPL EGAYDPADYE HLPVSAEIKE LFQYISRYTP QLIDLHKLK
101 PFIPDFIPAV GIDDAFLKVP RPDGKPDNLG LLVLDEPSTK QSDPTVLSLW
151 LTENSKQHNI TQHMVKVSL EAEKNPKAID TWIESISELH RSKPPATVHY
201 TRPMPDIDTL MQEWSPEFEE LLGKVSLEPTA EIDCSLAEYI DMICAILDIP
251 VYKSRIQSLH LLFSLYSEFK NSQHFALAE GKKAFTPSSN STSQAGDMET
301 LTFS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_22n2, frame 3

PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*), N = 1,
Score = 132, P = 1e-05

>PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*)
Length = 562

HSPs:

Score = 132 (19.8 bits), Expect = 1.0e-05, P = 1.0e-05
Identities = 24/63 (38%), Positives = 35/63 (55%)

```

Query:      3  DNSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSSSETDSDSDDDEEHGAPLEG 62
              +  DE EEE++ E++ T          +++DDDDDDDD + D D DDD++E A  G
Sbjct:     497 EEDDDDEDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDEDEDEAETPG 556

Query:      63 AYD 65
              D
Sbjct:     557 IID 559

```

Score = 122 (18.3 bits), Expect = 1.4e-04, P = 1.4e-04
Identities = 20/52 (38%), Positives = 33/52 (63%)

```

Query:      4  NSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSSSETDSDSDDDEE 55
              N+ +E ++E+ +E      + T + + N+DDDDDDDD + D D DDDD++
Sbjct:     494 NNEEEDDDDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDDDD 545

```

Pedant information for DKFZphut1_22n2, frame 3

Report for DKFZphut1_22n2.3

```

[LENGTH]      304
[MW]           34285.85
[pI]           4.37
[PROSITE]      AMIDATION      1
[PROSITE]      CAMP_PHOSPHO_SITE  2
[PROSITE]      CK2_PHOSPHO_SITE   10
[PROSITE]      PKC_PHOSPHO_SITE   1
[PROSITE]      ASN_GLYCOSYLATION  3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY     11.84 %

```

```

SEQ  MADNSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSSSETDSDSDDDEEHGAPL
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  cccccchhhhhchhhhhhhcccccccccccccccccccccccccccccccccccccc

```

```

SEQ  EGAYDPADYEHLFVSAEIKELFQYISRYTPQLIDLHKLKPFIPDFIPAVGDIDAFKVP
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhhhccccccccccccccccccccccccceec

```

```

SEQ  RPDGKPDNLGLLVLEPSTKQSDPTVLSLWLTENSKQHNIQHMVKVSLAEAEKNPKAID
SEG  .....
PRD  cccccccccceccccccccccccchhhhhhhccccccccccccchhhhhhhccccch

```

```

SEQ  TWIESISELHRSKPPATVHYTRPMPDIDTLMQEWSPEFEELGKVSLEPTAEIDCSLAEYI
SEG  .....
PRD  hhhhhhhhhccccceccccccccchhhhhhhccccchhhhhccccccccccccchhhhhh

```

```

SEQ  DMICAILDIPVYKSRIQSLHLLFSLYSEFKNSQHFALAEKGKKAFTPSSNSTSQAGDMET
SEG  .....

```

```
PRD      hhhhhhhhccccchhhhhhhhhhhhhhhhhhhcchhhhhhhccccc
```

```
SEQ      LTFS
```

```
SEG      . . .
```

```
PRD      cccc
```

Prosites for DKFZphute1_22n2.3

PS000001	4->8	ASN_GLYCOSYLATION	PDOC000001
PS000001	159->163	ASN_GLYCOSYLATION	PDOC000001
PS000001	290->294	ASN_GLYCOSYLATION	PDOC000001
PS000004	17->21	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	18->22	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	138->141	PKC_PHOSPHO_SITE	PDOC000005
PS000006	5->9	CK2_PHOSPHO_SITE	PDOC000006
PS000006	30->34	CK2_PHOSPHO_SITE	PDOC000006
PS000006	43->47	CK2_PHOSPHO_SITE	PDOC000006
PS000006	45->49	CK2_PHOSPHO_SITE	PDOC000006
PS000006	47->51	CK2_PHOSPHO_SITE	PDOC000006
PS000006	49->53	CK2_PHOSPHO_SITE	PDOC000006
PS000006	168->172	CK2_PHOSPHO_SITE	PDOC000006
PS000006	181->185	CK2_PHOSPHO_SITE	PDOC000006
PS000006	185->189	CK2_PHOSPHO_SITE	PDOC000006
PS000006	235->239	CK2_PHOSPHO_SITE	PDOC000006
PS000009	280->284	AMIDATION	PDOC000009

(No Pfam data available for DKFZphut1_22n2.3)

DKFZphut1_22o2

group: uterus derived

DKFZphut1_22o2 encodes a novel 537 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.pombe SPBC3E7.03c

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: map="11p15.5"

Insert length: 2714 bp

Poly A stretch at pos. 2695, polyadenylation signal at pos. 2677

```
1 GCAGGGCAGC GTGGGGGCTG AGATCGTTTC CTGTTGGAAC TTCTGGCCCA
51 AGAAGCGCGG GTCACAAGGA GAGGGGTCAG TTCGGTTCAG AGCGACTCAG
101 CCCCTCGACT CGGGTCTTAA AACCTCCGAG CCGCCAGTTC TGCTCAGGCG
151 CGCGCCCCCT TAAAGCGCCA CCAGACGCTG CGCCCCGTTA AAGCGCCACC
201 AGACGCCGCG CCCCGTCCCG GCCTCCCCCG CGCGCTGGCG CGGGGCTTTC
251 TGGGGCAGGG CGGGGCCGGC GAACTGCGGC CCGGAACGGC TGAGGAAGGG
301 CCCGTCCCGC CTTCCCGGGC GCGCCATGGA GCCCGGGGCG GTTGCAGAAG
351 CCGTGGAGAC GGGTGAGGAG GATGTGATTA TGGAACTCTT GCGGTCATAC
401 AACCAGGAGC ACTCCCAGAG CTTACGTTT GATGATGCCC AACAGGAGGA
451 CCGGAAGAGA CTGGCGGAGC TGCTGGTCTC CGTCTGGAA CAGGGCTTGC
501 CACCTCCCA CCGTGTCAATC TGGCTGCAGA GTGTCCGAAT CCTGTCCCGG
551 GACCCCAACT GCCTGGACCC GTTACCAGC CGCCAGAGCC TGCAGGCACT
601 AGCCTGCTAT GCTGACATCT CTGTCTCTGA GGGGTCCGTC CCAGAGTCCG
651 CAGACATGGA TGTGTACTG GAGTCCCTCA AGTGCTGTG CAACCTCGTG
701 CTCAGCAGCC CTGTGGCACA GATGCTGGCA GCAGAGGCC GCCTAGTGGT
751 GAAGCTCACA GAGCGTGTGG GGCTGTACCG TGAGAGGAGC TTCCCCCAGC
801 ATGTCCAGTT CTTTGACTTG CGGCTCCTCT TCCTGTCTAA GGCACCTCCG
851 ACCGATGTGC GCCAGCAGCT GTTTCAGGAG CTGAAAGGAG TCGGCCTGCT
901 AACTGACACA CTGGAGCTGA CGCTGGGGGT GACTCCTGAA GGAACCCCC
951 CACCCACGCT CTTTCTCTCC CAAGAGACTG AGCGGGCCAT GGAGATCCTC
1001 AAAGTGCTCT TCAACATCAC CCTGGACTCC ATCAAGGGGG AGGTGGACGA
1051 GGAAGACGCT GCCCTTACC GACACCTGGG GACCTTCTC CGGCACTGTG
1101 TGATGATCGC TACTGTCTGA GACCGCACAG AGGAGTTCCA CGGCCACGCA
1151 GTGAACCTCC TGGGGAACCT GCCCTCAAG TGTCTGGATG TTCTCCTCAC
1201 CTTGAGGCCA CATGGAGACT CCACGGAGTT CATGGGAGTG AATATGGATG
1251 TGATTCGTGC CTTCTCATC TTCCTAGAGA AGCGTTTGCA CAAGACACAC
1301 AGGCTGAAGG AGAGTGTAGC TCCCGTGTG AGCGTGTCTA CTGAATGTGC
1351 CCGGATGCAC CGCCAGCCA GGAAGTTCCT GAAGGCCAG GGATGGCCAC
1401 CTCCCCAGGT GTCGCCCTC CTGCGGGATG TGAGGACACG CCCTGAGGTT
1451 GGGGAGATGC TGGGAACAA GCTTGTCGCG CTCATGACAC ACCTGGACAC
1501 AGATGTGAAG AGGGTGGCTG CCGAGTCTTT GTTGTCTCTG TGCTCTGAGA
1551 GTGTGCCCGG ATTCAACAAG TACACAGGCT ATGGGAATGC TGCTGGCCTT
1601 CTGGCTGCCA GGGCCCTCAT GGCAGGAGGC CGGCCGAGG GCCAGTACTC
1651 AGAGGATGAG GACACAGACA CAGATGAGTA CAAGGAAGCC AAAGCCAGCA
1701 TAAACCTGTG GACCGGAGG GTGGAGGAGA AGCCGCTTAA CCCTATGGAG
1751 GGCATGACAG AGGAGCAGAA GGAGCACGAG GCCATGAAGC TGGTGACCAT
1801 GTTTGACAAG CTCTCCAGGA ACAGAGTCAT CCAGCCAATG GGGATGAGTC
1851 CCCGGGGTCA TCTTACGTCC CTGCAGGATG CCATGTGCGA GACTATGGAG
1901 CAGCAGCTCT CTTCCGACCC TGAATCGGAC CCTGACTGAG GATGGCAGCT
1951 CTTCTGCTCC CCCATCAGGA CTGGTGTCTG TTCCAGAGAC TTCCTTGGGG
2001 TTGCAACCTG GGAAGGCCAC ATCCCCTGAG ATCCACACCC GCCCCCACTT
2051 CTCCATCTTA GAAACCCCTT CTCTTGACTC CCGTTCTGTT CATGATTGTC
2101 CTCTGGTCCA GTTCTCATC TCTGGACTGC AACGGTCTTC TTGTGCTAGA
2151 ACTCAGGCTC AGCCTCGAAT TCCACAGACG AAGTACTTTC TTTTGTCTGC
2201 GCCAAGAGGA ATGTGTTCAG AAGCTGCTGC CTGAGGGCAG GGCTACCTTG
2251 GGCACACAGA AGAGCATATG GGAGGGCAGG GGTTTGGGTG TGGGTGCACA
2301 CAAAGCAAGC ACCATCTGGG ATTGGCACAC TGGCAGAGCC AGTGTGTTGG
2351 GGTATGTGCT GCACTTCCCA GGGAGAAAAC CTGTCAAGAC TTTCATACG
2401 AGTATATCAG AACACACCTT TCCAAGGTAT GTATGCTCTG TTGTTCCTGT
2451 CCTGTCTTCA CTGAGCGCAG GGCTGGAGGC CTCTTAGACA TTCTCCTTGG
2501 TCCTCGTTCA GCTGCCCACT GTAGTATCCA CAGTGCCCGA GTTCTCGCTG
2551 GTTTTGGCAA TTAAACCTCC TTCCTACTGG TTAGACTAC ACTTACAACA
2601 AGGAAATGCG CCTCGTGTG ACCATAGATT GAGATTATTA CCACATACCA
2651 CACATAGCCA CAGAAACATC ATCTTGAAAT AAAGAAGAGT TTTGGACAAA
2701 AAAAAAAAAA AAAA
```


BLAST Results

Entry AF015416 from database EMBL:
Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.
Score = 3356, P = 2.0e-144, identities = 672/673

Entry HS263253 from database EMBL:
human STS SHGC-15914.
Score = 1143, P = 9.0e-46, identities = 245/255

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 326 bp to 1936 bp; peptide length: 537
Category: similarity to unknown protein

```

1 MEPRVAEAV ETGEEDVIME ALRSYNQEH SFTFDDAQQ EDRKRLAELL
51 VSVLEQGLPP SHRVIWLOS RILSRDRNCL DPFTSRQSLQ ALACYADISV
101 SEGVPESAD MDVVLES LKCNLVLSSPV AQMLAAEARL VVKLTERVGL
151 YRERSFPHDV QFFDLRLFL LTALRTDVRQ QLFQELKGV RLLDTLELTL
201 GVTPEGNPPP TLLPSQETER AMEILKVLFN ITLDSIKGEV DEEDAALYRH
251 LGTLLRHCV M IATAGDRTEE FHGHAVNLLG NLPLKCLDVL LTLEPHGDST
301 EFMGVNMDVI RALLIFLEKR LHKTHRLKES VAPVLSVLTE CARMHRPARK
351 FLKAQGWPPP QVLPPLRDVR TRPEVGEMLR NKLVRMLTHL DTDVKRVAE
401 FLFVLCSESV PRFIKYTG YG NAAGLLAARG LMAGGRPEGQ YSEDEDTDT
451 EYKEAKASIN PVTGRVEEK PNPMEGMTEE QKEHEAMKL VTMFDKLSRNR
501 VIQPMGMSPR GHLTSLQDAM CETMEQQLSS DPDSDPD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phutell_22o2, frame 2

TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7., N = 1, Score = 112, P = 0.0023

>TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7.
Length = 362

HSPs:

Score = 112 (16.8 bits), Expect = 2.3e-03, P = 2.3e-03
Identities = 71/289 (24%), Positives = 124/289 (42%)

```

Query: 215 SQETERAM-EILKVLFNITLDSIKGEVDEEDAALYRHLGTLRLHCVMIATAGDRTEEFHG 273
      SQ+ E + EIL++LF I+ S E DE+ L L+ + +
Sbjct: 12 SQDNEMVLTEILRLFLPISKRSYLKEEDEQKILL-----LVIEIWASSLNNNPNSPLRW 65

Query: 274 HAVN-LLG-NLPLKCLDVLTLTLEPHGDSTEFMGVNMDVIRALLIFLEKRLHKTH----RL 327
      HA N LL NL L LD + + T + +I + +LEK L+ +
Sbjct: 66 HATNALLSFNLQLLSLDQAIYVSEIACQT----LQSILISREVEYLEKGLNLCFDIAAKY 121

Query: 328 KESVAPVLSVLTECARMHRPARKFLKAQGWPPPQVLPPLRDVTRTP-EVGEMLRNKLVR 386
      + ++ P+L++L + +L P D R + + G+ R L+RL
Sbjct: 122 QNTLPPIAILLSLLSFFNIKQNL-----SMLLFTNDDRKQSLQKGSFRCLLLRL 173

Query: 387 MT-HLDTDVKRVAEFLFVLCSESVPRFIKYTGYNAAAGLLAARGLMAGGRPEGQYS--- 442
      +T + + A L LC + + G G A G+ M P + +
Sbjct: 174 LTIPIVEPIGTYYASLLNELCDGDSQQIARIFGAGYAMGISQHSETMPFPSPLSKAASPV 233

Query: 443 -EDEDTDDEYKEAKASINPVTGRV--EEKPPNPMEGMTEEQKEHEAMKLVTMFDKLSRN 499
      + + +E +I+P+TG + +E +++E+KE EA +L +F +L +N
Sbjct: 234 FQKNSRGQENTEENLAIDPITGSMCTNRNKSQRLE-LSQEEKEREERLFFYLFQRLEKN 292

```

Pedant information for DKFZphutel 22o2, frame 2

Report for DKFZphutel 22o2.2

Prosites for DKFZphute1 22o2.2

523

PS00006	388->392	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	530->534	CK2_PHOSPHO_SITE	PDOC00006
PS00008	57->63	MYRISTYL	PDOC00008
PS00008	420->426	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	430->436	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphute1_22o2.2)

DKFZphut1_23e13

group: metabolism

DKFZphutes3_15j18 encodes a novel 148 amino acid protein with similarity to 27K heat shock proteins.

The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases. Thus the novel protein is a new member of this family.

The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnologic production processes.

heat shock protein HSP27

strong similarity to heat shock 27K proteins

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: /map="578.9 cR from top of Chr12 linkage group"

Insert length: 1854 bp

Poly A stretch at pos. 1831, polyadenylation signal at pos. 1810

```
1  GGGTTATTAA  GCTCCTGGCT  CCGCTCTAGA  CCTCAGCGGT  TCTGGCTGCC
51  AGCCTGGGCA  GCCTGGGAAG  CCTGGGAGGA  CGGTGGCTTG  CCGGTCTGTC
101  GTGAGGCAGT  GCGGACGGGG  ACCCTCTGGG  ATTCTGCTGG  ATCTGCCCCG
151  GGGGTACCTT  TTGGGGGCTG  GGACCCAGT  CGAGGGGACA  CAACCGTCCC
201  TGGCAGTGGT  TGGTTCTGCT  TCTCCCTGCA  GAAAAGCAGC  ATTTTCGGAA
251  GCTGAAGAA  AAGCTAGCCC  AGCCACACCA  CCTTGTGTG  TGACCTTGGG
301  CAGGTGGTTC  TGTCTCTCTG  AGCCTCTGTT  TCTCTCTGAG  CTGAGCAGCC
351  ACCATGGCTG  ACGGTCAGAT  GCCCTTCTCC  TGCCACTACC  CAAGCCGCCT
401  GCGCCGAGAC  CCCTTCCGGG  ACTCTCCCT  CTCCTCTCGC  CTGCTGGATG
451  ATGGCTTTGG  CATGGACCCC  TTCCAGACG  ACTTGACAGC  CTCTTGGCCC
501  GACTGGGCTC  TGCCTCGTCT  CTCCTCCGCC  TGGCCAGGCA  CCCTAAGGTC
551  GGGCATGGTG  CCCCAGGGCC  CCACTGCCAC  CGCCAGGTTT  GGGGTGCTG
601  CCGAGGGCAG  GACCCCCCA  CCCTTCCCTG  GGGAGCCCTG  GAAAGTGTGT
651  GTGAATGTGC  ACAGCTTCAA  GCCAGAGGAG  TTGATGGTGA  AGACCAAAGA
701  TGGATACGTG  GAGGTGTCTG  GCAAACATGA  AGAGAAACAG  CAAGAAGGTG
751  GCATTGTTTC  TAAGAACTTC  ACAAGAAAA  TCCAGCTTCC  TGCAGAGGTG
801  GATCCTGTGA  CAGTATTGTC  CTCCTTTCC  CCAGAGGGTC  TGCTGATCAT
851  CGAAGCTCCC  CAGGTCCCTC  CTTACTCAAC  ATTTGGAGAG  AGCAGTTTCA
901  ACAAGAGAGT  TCCCCAGGAC  AGCCAGGAAG  TCACCTGTAC  CTGAGATGCC
951  AGTACTGGCC  CATCCTTGTT  TTGTCCCAA  CCCTAGGGCT  TCTCTGATTC
1001  CAGGATACAT  TACTTTAGCT  GAACTCAGAT  TTAGTGCAAG  TAAATGTGTA
1051  GAGGTGCGG  GGGTGAGGAC  TGACCACAGA  TTCCCTGGAT  AGTGTAGTGG
1101  TAGATTTCTC  CACAGGATAG  CGCAATTGGC  AAATCATGCT  TGTTGTGTT
1151  AGGCCAAAT  ACTAGTTTG  CTTTCTTAC  CTTTCTATC  TTGATGAAAA
1201  TGTGACACAT  TCTATAGTTG  CAAAACACAT  AAAAGGGGAC  TTAACATTTC
1251  ACGTTGTATC  TTAATTGCAG  TGAATGCAAG  GGTACTTTT  CTCTGGGGAC
1301  CTCCCCATC  ACCCAGGTT  CTAATCTGGG  CTCCCGATTC  CCATGGCTCC
1351  CAAACCATGC  CGCATGGTTT  GGTTAATGAA  ACCCAGTAGC  TAACCCCACT
1401  GTGCTTCCAC  ATGCCTGGCC  TAAATGGGT  GATATACAGG  TCTTATATCC
1451  CCATATGGAA  TTTATCCATC  AACCACATAA  AAACAACAG  TGCCCTTCTG
1501  CTTCTGCCCA  GATGTGTCCA  GCACGTTCTC  AAAGTTTCCA  CATTAGCACT
1551  CCCTAAGGAC  GCTGGGAGCC  TGTCACTTTA  TGATCTGACC  TAGGTCCCCC
1601  CTTTCTTCTG  TCCCCTGTGT  TTAAGTCGGG  ATTTTACAG  AGGGAGCTGT
1651  CTCCAGACAG  CTCCATCAG  AACCAAGCAA  AGGCCAGATA  GCCTGACAGA
1701  TAGGCTAGTG  GTATTGTGTA  TATGGGCGGG  ACGTGTGTGT  CATTATTATT
1751  TGAGTTATGC  TGTGTGTTAG  GGGTAAATAA  CAGTAAATAA  TTAATAATAA
1801  TAATAATAAT  AATAAAGGAG  CTGACGTTCT  TAAAAAGGAA  AAAAAAATAA
1851  AAAA
```

BLAST Results

Entry HS286348 from database EMBL:

human STS TIGR-A002J47.

Score = 510, P = 1.2e-16, identities = 102/102

Medline entries

95394379:
Cloning and sequencing of a cDNA encoding the canine HSP27 protein.

94110260:
Physiological and pathological changes in levels of the two small stress proteins, HSP27 and alpha B crystallin, in rat hindlimb muscles

Peptide information for frame 3

ORF from 354 bp to 941 bp; peptide length: 196
Category: strong similarity to known protein
Prosites motifs: SUBTILASE_ASP (28-39)

1 MADGQMPFSC HYP SRLRRDP FRD SPLSSRL LDDGFGMDPF PDDL TASWPD
51 WALPRLSSAW PGT LRSGMVP RGPTATARFG VPAEGRTPPP FPGEPWKVCV
101 NVHSFKPEEL MVKTKDGYVE VSGKHEEKQQ EGGIVSKNFT KKIQLPAEVD
151 PVTVFASLSP EGLLIIEAPQ VPPYSTFGES SFNNELPQDS QEVTCT

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_23e13, frame 3

PIR:JC4244 heat-shock 27K protein - dog, N = 1, Score = 304, P = 4.3e-27

PIR:JN0924 heat shock 27 protein - rat, N = 1, Score = 301, P = 8.9e-27

TREMBL:MM03561_1 product: "heat shock protein HSP27"; Mus musculus heat shock protein HSP27 internal deletion variant b mRNA, complete cds., N = 1, Score = 301, P = 8.9e-27

>PIR:JC4244 heat-shock 27K protein - dog
Length = 209

HSPs:

Score = 304 (45.6 bits), Expect = 4.3e-27, P = 4.3e-27
Identities = 80/182 (43%), Positives = 102/182 (56%)

Query: 1 MADGQMPFSC-HYP SRLRRDPFRD-SPLSSRLDDGFGMDPFDDL TASWPDWALPRLSS 58
M + ++PFS PS DPFRD P SRL D FG+ P++ W W S
Sbjct: 1 MTERRVPFSLRSPSW---DPFRDWYPAHSRLFDQAFLPRLPEE----WAQWFG---HS 50

Query: 59 AWPGLRSGMVP---RGPTATARFGVPAEGR--TPPPFFG-----EPWKVCVNVHSF 105
WPG +R +P GP A A PA R + G + W+V ++V+ F
Sbjct: 51 GWPGYVRP--IPPAVEGPAAAAAAPAYSRLSRQLSSGVSEIRQTADRWRVSLDVNHF 108

Query: 106 KPEELMVKTKDGYVEVSGKHEEKQQEGGIVSKNFTKKIQLPAEVD PVTVFASLSPEGLLI 165
PEEL VKTKDG VE++GKHEE+Q E G +S+ T K LP VDP V +SLSPEG L
Sbjct: 109 APEELTVKTKDGVVEITGKHEERQDEHGYISRRLLTPKYTLPPGVDP TLVSSSLSP EGTLT 168

Query: 166 IEAPQVPPYSTFGE 179
+EAP P + E
Sbjct: 169 VEAPMPKPATQSAE 182

Pedant information for DKFZphut1_23e13, frame 3

Report for DKFZphut1_23e13.3

[LENGTH] 196
[MW] 21604.37

[pI] 5.00
 [HOMOL] PIR:JC4244 heat-shock 27K protein - dog 3e-22
 [BLOCKS] BL01031C
 [PIRKW] blocked amino end 1e-13
 [PIRKW] acetylated amino end 4e-13
 [PIRKW] phosphoprotein 7e-21
 [PIRKW] glycoprotein 2e-11
 [PIRKW] heat shock 7e-21
 [PIRKW] molecular chaperone 4e-13
 [PIRKW] alternative splicing 1e-19
 [PIRKW] eye lens 6e-14
 [PIRKW] stress-induced protein 7e-21
 [SUPFAM] alpha-crystallin 7e-21
 [PROSITE] SUBTILASE_ASP 1
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 6
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Heat shock hsp20 proteins
 [KW] All_Beta
 [KW] LOW_COMPLEXITY 7.14 %

SEQ MADGQMPPFSCHYPSRLRRDPFRDSPLSSRLDDGFGMDPFPDDLTA SWPDWALPRLSSAW
 SEGXXXXXXXXXXXXXXXXX.....
 PRD cccccccccccccccccccccccccchhhhhcccccccccccccccccccccccccccc

SEQ PGTLRSGMVPRGPTATARFGVPAEGRTPPFPFGEPPWKVCNVHSFKPEELMVKT KDGYVE
 SEG
 PRD cccccccccccccchhhhhhhccccccccchhhhhheeeeeccccceeeeeeccccccc

SEQ VSGKHEEKQEGGIVSKNFTKKIQLPAEVDPVTVFASLSPEGLLIIEAPQVPPYSTFGES
 SEG
 PRD eccchhhhhccccceeeccccccccccccccccceeeccccceeeccccccccccccccc

SEQ SFNNELPQDSQEVCTCT
 SEG
 PRD cccccccccceeeccc

Prosites for DKFZphut1_23e13.3

PS00001	138->142	ASN_GLYCOSYLATION	PDOC00001
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	140->143	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	62->68	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00136	28->39	SUBTILASE_ASP	PDOC00125

Pfam for DKFZphut1_23e13.3

HMM_NAME	Heat shock hsp20 proteins		
HMM	*AMMrpPQDWRE....DpDHFeVrMDMPGFKPEEIKVkvEDNNVLvIeG		
	A	P++ R	+ ++V+++ FKPEE+ VK+ D+ ++++G
Query	77	AREGVPAEGR-TPPFPFGEPPWKVCNVHSFKPEELMVKT KDGYVEVSG	123
HMM	EHEREEREDDkWWHERIYRHFMRFRrLPENVDpDqIkAsMSdNGVLT I		
	+HE	E++	+ + ++ F +++LP +VDP + AS+S++G+L I
Query	124	KHE---EKQQ---EGGIVSKNFTKKIQLPAEVDPVTVFASLSPEGLLI I	166
HMM	TVPKpEP*		
	++P ++P		
Query	167	EAPQVPP	173

DKF2phut1_23g11

group: uterus derived

DKF2phut1_23g11 encodes a novel 256 amino acid protein with similarity to *S.pombe* SPAC31G5.12c and *S. cerevisiae* Maf1p.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to SPAC31G5.12c and Maf1p

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1674 bp

Poly A stretch at pos. 1664, polyadenylation signal at pos. 1644

```
1  GGGGGAGGCG GAGGTCGCTC GCTCGCTCGC TCGGCTCGCT GACTCGCCGG
51 AGCGCTCTGT GGCGGTCGGC GGCAGGTCGG TCGCGAGAGC GGGCTCTGTG
101 GAAGGGGGCG AGGCTATGTC GCGGTGGCAG CCCGGATGGG CCGGCAGGGC
151 CGGGAGTAAC GGGACGTCGC CGCGGAGCTT CTTCCCCCGG ATACAGTGCG
201 GCGCGAGCGG AGGCCGCGGC GCCGCCCTCC GATCTTGAAG AGCCCCGCGT
251 GCGCGGAGCC CGCCCCCGCC TCGGCACCGG CACCGACGCG GAGCGACCCAG
301 CCCAGCCAGA CCCGGCCCGG CGCGGCCTGA TCTAACCAG CCAGGCAGGC
351 AATACTAGCC CCTCTGGAGC ACGGAGCTCC TTCCCCAAG ACATGAAGCT
401 ATTTGAGAAC TCGAGCTTTG AAGCCATCAA CTCACAGCTG ACTGTGGAGA
451 CCGGAGATGC CCACATCATT GGCAGGATTG AGAGCTACTC ATGTAAGATG
501 GCAGGAGACG ACAAAACACAT GTTCAAGCAG TTCTGCCAGG AGGGCCAGCC
551 CCACGTGCTG GAGGCACCTT CTCCACCCCA GACTTCAGGA CTGAGCCCCA
601 GCAGACTCAG CAAAAGCCAA GCGGTTGAGG AGGAGGGGCC CCTCAGTGAC
651 AAGTGCAGCC GCAAGACCCT CTTCTACCTG ATTGCCACGC TCAATGAGTC
701 CTTCAGGCCT GACTATGACT TCAGCACAGC CCGCAGCCAT GAGTTCAGCC
751 GGGAGCCGAG CCTTAGCTGG GTGGTGAATG CAGTCAACTG CAGTCTGTTT
801 TCAGCTGTGC GGGAGGACTT CAAGGATCTG AAACCACAGC TGTGGAACGC
851 GGTGGACGAG GAGATCTGCC TGGCTGAATG TGACATCTAC AGCTATAACC
901 CAGACTTGGA CTCAGATCCC TTCGGGGAGG ATGGTAGCCT CTGGTCCTTC
951 AACTACTTCT TCTACAACAA GCGGCTCAAG CGAATCGTCT TCTTTAGCTG
1001 CCGTTCATC AGTGGCTCCA CCTACACACC CTCAGAGGCA GGCAACGAGC
1051 TGGACATGGA GCTGGGGGAG GAGGAGGTGG AGGAAGAAAG CAGAAGCAGG
1101 GGCAGTGGGG CCGAGGAGAC CAGCACCATG GAGGAGGACA GGGTCCCAGT
1151 GATCTGTATT TGATGAGGAG GAGCCGAGGC CCCAGCTTCA TCCAGCTTCA
1201 ACCAATGCCT GGACCTGTCC ACCTGAGAGG CCCCTGGGGC CTCCCCAGCT
1251 GCTGGCCAGA CCCTGGCGCT GCCACAGTCC TGGCACTGCC CAAGGCCATA
1301 CCTGCCTAGC CTTTGGCTC CATCCTGTGG ATGCCCACTC ACCCCTCAGA
1351 CTCCTGCTGC CCATGCTGTG GCCGGAATG TCAGCAGGGG GCCTGGTGGG
1401 AGGAGCGACT GCCCTGCCCA AATGAACTGC CACAGCAGGG ACAGCTGGAC
1451 CGCAGAGTTT ATTTTGTAT TTCTACTGGG CTGCACACT CCAGCCCAAA
1501 GGGTCTGTGG CCGGAGGCC CACGAGCAGG CCCCAGCAGT CACCGGCTCT
1551 GGTCTTGGGC CGGCCCGGT GCCCACCTGT ACCCCACCT CGCCCATTTG
1601 GCCGGGTGCA CTGAGTGTC CTTGTGTCGA GCTCGTTTCT TTCCAATAAA
1651 AGTTTCTGTG ACTTAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 393 bp to 1160 bp; peptide length: 256

Category: similarity to known protein

BLASTP hits

Report for DKFZphutel 23g11.3

529

Prosites for DKFZphut1_23g11.3

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00001	132->136	ASN_GLYCOSYLATION	PDOC00001
PS00005	33->36	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	181->187	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_23g11.3)

DKFZphut1_24c19

group: transmembrane protein

DKFZphut1_24c19 encodes a novel 195 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

unknown

membrane regions: 1

Summary DKFZphut1_24c19 encodes a novel 195 amino acid protein, with no similarity to known proteins.

unknown

complete cDNA, complete cds, EST hits

TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 769 bp

Poly A stretch at pos. 746, polyadenylation signal at pos. 735

```

1 ACGAGTCAGC CAAAGATGGC TGCGCCAGG TAATTGAGC AAAGGCCACA
51 GTGAACCTCG GCGTGGCTGA GGAAGACCGG AGGAGGCACC CACAGGCTGC
101 TGGGAGGAGA GCATAAGGCT CAAAATGGAA AATCATAAAT CCAATAATAA
151 GGAAACATA ACAATTGTTG ATATATCCAG AAAAATTAAC CAGCTTCCAG
201 AAGCAGAAAG GAATCTACTT GAAAATGGAT CGGTTTATGT TGGATTAAAT
251 GCTGCTCTTT GTGGCCTCAT AGCAAACAGT CTTTTTCGAC GCATCTTGAA
301 TGTGACAAAG GCTCGCATAG CTGCTGGCTT ACCAATGGCA GGGATACCTT
351 TTCTTACAAC AGACTTAACT TACAGATGTT TTGTAAGTTT TCCTTTGAAT
401 ACAGGTGATT TGGATTGTGA AACCTGTACC ATAACACGGA GTGGACTGAC
451 TGGTCTTGTT ATTGGTGGTC TATACCCTGT TTTCTGGCT ATACCTGTAA
501 ATGGTGGTCT AGCAGCCAGG TATCAATCAG CTCTGTTACC ACACAAAGGG
551 AACATCTTAA GTTACTGGAT TAGAACTTCT AAGCCTGTCT TTAGAAAGAT
601 GTTATTTCTT ATTTTGCTCC AGACTATGTT TTCAGCATAC CTTGGGCTGT
651 AACAAATATA ACTACTTATA AAGGCCCTTC AGTTATCTGA ACCTGGCAAA
701 GAAATTCAC TATTTTAAAC AAATATGTAA ACAAATAA AATGGTAAAA
751 ACAAAAAAAA AAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 709 bp; peptide length: 195

Category: putative protein

```

1 MENHKSNNKE NITIVDISRK INQLPEAERN LLENGSVYVG LNAALCGLIA
51 NSLFRRILNV TKARIAAGLP MAGIPFLTDD LTYRCFVSFP LNTGDLDCET
101 CTITRSLTGT LVIGGLYPVF LAIPVNGGLA ARYQSALLPH KGNILSYWIR
151 TSKPVFRKML FPILLQTMFS AYLGSEQYKL LIKALQLSEP GKEIH

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24c19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut1_24c19, frame 2

Report for DKFZphut1_24c19.2

```
[LENGTH]      195
[MW]           21527.45
[pI]           9.36
[PROSITE]      MYRISTYL      6
[PROSITE]      CK2_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      3
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           TRANSMEMBRANE 1

SEQ  MENHKSNNKENITIVDISRKINQLPEARNLLENGSVYVGLNAALCGLIANSLFRRILNV
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  TKARIAAGLPAGIPFLTTDLTYRCFVSFPLNTGDLDCETCTITRSGLTGLVIGGLYPVF
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMM

SEQ  LAIPVNGGLAARYQSALLPHKGNILSYWIRTSKPVFRKMLFPILLQTMFSAYLGSEQYKL
PRD  eeccccccchhhhhcccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMM.....

SEQ  LIKALQLSEPGKEIH
PRD  hhhhhhcccccccc
MEM  .....
```

Prosites for DKFZphut1_24c19.2

PS00001	11->15	ASN_GLYCOSYLATION	PDOC00001
PS00001	34->38	ASN_GLYCOSYLATION	PDOC00001
PS00001	59->63	ASN_GLYCOSYLATION	PDOC00001
PS00005	18->21	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	151->154	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00008	40->46	MYRISTYL	PDOC00008
PS00008	47->53	MYRISTYL	PDOC00008
PS00008	68->74	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	142->148	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_24c19.2)

DKFZphut1_24e11

group: intracellular transport and trafficking

DKFZphut1_24e11 encodes a novel 226 amino acid protein, with similarity to human/mouse golgi 4-transmembrane spanning transporter MTP. MTP may function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment. Thus, the novel protein also seems to be involved in nucleotide sugar transport.

The new protein can find application in modulating the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartments.

similarity to 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP

complete cDNA, complete cds, EST hits

potential start at 184,

TRANSMEMBRANE 4

function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment?

Sequenced by Qiagen

Locus: /map="8"

Insert length: 2005 bp

Poly A stretch at pos. 1988, polyadenylation signal at pos. 1963

```
1  ACGCGTCCGG CAGAAGCTCG GAGCTCTCGG GGTATCGAGG AGGCAGGCC
51  GCGGGCGCAC GGGCGAGCGG GCCGGGAGCC GGAGCGCGG AGGAGCCGGC
101 AGCAGCGGGG CGGCGGGCTC CAGGCGAGGC GGTGACGCT CCTGAAACT
151 TGCGCGCGCG CTCGCGCCAC TCGCGCCGGA GCGATGAAGA TGGTCGCGCC
201 CTGGACGCGG TTCTACTCCA ACAGCTGCTG CTTGTGCTGC CATGTCCGCA
251 CCGGACCATC CTGCTCGGCG GTCTGGTATC TGATCATCAA TGCTGTGGTA
301 CTGTTGATTT TATTGAGTGC CCTGGCTGAT CCGGATCAGT ATAACCTTTC
351 AAGTTCTGAA CTGGGAGGTG ACTTTGAGTT CATGGATGAT GCCAACATGT
401 GCATTGCCAT TCGGATTTCT CTCTCATGA TCCTGATATG TGCTATGGCT
451 ACTTACGGAG CGTACAAGCA ACGCGCAGCC TGGATCATCC CATTCTTCTG
501 TTACCAGATC TTTGACTTTG CCCTGAACAT GTTGGTTGCA ATCAGTGTGC
551 TTATTATATC AAACCTCCAT CAGGAATACA TACGGCAACT GCCTCCTAAT
601 TTTCCTTACA GAGATGATGT CATGTCAGTG AATCCTACCT GTTTGGTCTC
651 TATTATCTTT CTGTTTATTA GCATTATCTT GACTTTTAAG GGTACTTTGA
701 TTAGCTGTGT TTGGAAGTGC TACCGATACA TCAATGGTAG GAACTCCTCT
751 GATGTCCTGG TTTATGTTAC CAGCAATGAC ACTACGGTGC TGCTACCCCC
801 GTATGATGAT GCCACTGTGA ATGGTGCTGC CAAGGAGCCA CCGCCACCTT
851 ACGTGCTGTC CTAAGCCTTC AAGTGGGCGG AGCTGAGGGC AGCAGCTTGA
901 CTTTGCAGAC ATCTGAGCAA TAGTTCTGTT ATTTCACTTT TGCCATGAGC
951 CTCTCTGAGC TTGTTTGTGT CTGAAATGCT ACTTTTAAAT ATTTAGATGT
1001 TAGATTGAAA ACTGTAGTTT TCAACATATG CTTTGCTAGA AACTGTGAT
1051 AGATTAACTG TAGAATCTTT CCTGTACGAT TGGGGATATA ACGGGCTTCA
1101 CTAACCTTCC CTAGGCATTG AAACCTCCCC CAAATCTGAT GGACCTAGAA
1151 GTCTGCTTTT GTACCTGCTG GGCCCCAAAG TTGGGCATTT TTCTCTCTGT
1201 TCCCTCTCTT TTGAAAATGT AAAATAAAAC CAAAAATAGA CAACTTTTTC
1251 TTCAGCCATT CCAGCATAGA GAACAAAACC TTATGGAAAC AGGAATGTCA
1301 ATTGTGTAAT CATTGTTCTA ATTAGGTAAA TAGAAGTCCT TATGTATGTG
1351 TTACAAGAAT TTCCCCACA ACATCCTTTA TGAAGTGAAG TCAATGACAG
1401 TTTGTGTTTG GTGGTAAAGG ATTTCTTCCA TGGCCTGAAT TAAGACCATT
1451 AGAAAGCACC AGGCCGTGGG AGCAGTGACC ATCTACTGAC TGTCTTGTG
1501 GATCTGTGTG CCAGGGACAT GGGGTGACAT GCCTCGTATG TGTTAGAGGG
1551 TGGAAATGGAT GTGTTTGGCG CTGCATGGGA TCTGGTCCCC CTCTTCTCCT
1601 GGATTACATC CCCCACCCAG GGCCCGCTTT TACTAAGTGT TCTGCCCTAG
1651 ATTGGTTCAA GGAGGTATC CAACTGACTT TATCAAGTGG AATTGGGATA
1701 TATTTGATAT ACTTCTGCCT AACACATGG AAAAGGGTTT TCTTTTCCCT
1751 GCAAGCTACA TCCTACTGCT TTGAACTTCC AAGTATGTCT AGTCACCTTT
1801 TAAAATGTAA ACATTTTCAG AAAAATGAGG ATTGCCTTCC TTGTATGCGC
1851 TTTTACCTTT GACTACCTGA ATTGCAAGGG ATTTTATAT ATTCATATGT
1901 TACAAAGTCA GCAACTCTCC TGTGTTGTTA TTATTGAATG TGCTGTAAAT
1951 TAAGTCGTTT GCAATTAAAA CAAGGTTTGC CCACATCCAA AAAAAA
2001 AAAAA
```

BLAST Results

Entry HS012351 from database EMBL:

human STS SHGC-31823.
Score = 1629, P = 3.1e-67, identities = 343/354

Medline entries

96199248:
Identification of a novel membrane transporter
associated with intracellular membranes by
phenotypic complementation in the yeast
Saccharomyces cerevisiae.

Peptide information for frame 1

ORF from 184 bp to 861 bp; peptide length: 226
Category: strong similarity to known protein

1 MKMVAPWTRF YNSCCCLCCH VRTGTILLGV WYLIINAVVL LILLSALADP
51 DQYNFSSSEL GGDFFEMDDA NMCIAIAISL LMILICAMAT YGAYKQRAAW
101 IIPFFCYQIF DFALNMLVAI TVLIYPNSIQ EYIRQLPPNF PYRDDVMSVN
151 PTCLVLIILL FISIIILTFKG YLISCVWNCY RYINGRNSSD VLVYVTSNDT
201 TVLLPPYDDA TVNGAAKEPP PPYVSA

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_24e11, frame 1

SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108)., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MTRP_MOUSE GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP., N
= 1, Score = 539, P = 5.3e-52

TREMBL:HS304981_1 product: "E3 protein"; Human retinoic acid-inducible
E3 protein mRNA, complete cds., N = 1, Score = 127, P = 3.4e-06

>SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
(KIAA0108).
Length = 233

HSPs:

Score = 551 (82.7 bits), Expect = 2.9e-53, P = 2.9e-53
Identities = 102/221 (46%), Positives = 148/221 (66%)

Query: 9 RFYSNSCCCLCHVRTGTILLGVWYLIINAVVL LILLSALADPDQY---NFSSELGGDF- 64
RFYS CC CCHVRTGTI+LG WY+++N ++ ++L + P+ N +G +
Sbjct: 13 RFYSTRCCGCCHVRTGTIILGTWYMVVLLMAILLTVEVTHPNSMPAVNIQYEVIGNYYS 72
Query: 65 -EFMDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVL 123
E M D N C+ A+S+LM +I +M YGA + W+IPFFCY++FDF L+ LVAI+ L
Sbjct: 73 SERMAD-NACVLFAVSVLMFISSMLVYGAIYQVGWLIIPFFCYRLFDVLSCLVAISSL 131
Query: 124 IYPNSIQEYIRQLPPNFYRDDVMSVNPCTCLVLIILLFISIIILTFKGYLISCVWNCYRYI 183
Y I+EY+ QLP +FPY+DD++++ +CL+ I+L+F ++ + FK YLI+CVWNCY+YI
Sbjct: 132 TYLPRIKEYLDQLP-DFPYKDDLLALDSSCLLFIVLVFFALFIIFKAYLINCWNCYKYI 190
Query: 184 NGRNSSDVLVYVTSN-DTTVLLPPYDDATVNGAAKEPPPPYVSA 226
N RN ++ VY +LP Y+ A V KEPPPPY+ A
Sbjct: 191 NNRNVPEIAVYPAFEAPPQYVLPTYEMA-VKMPEKEPPPPYLPA 233

Pedant information for DKFZphut1_24e11, frame 1

Report for DKFZphut1_24e11.1

[LENGTH] 226
[MW] 25419.11

[pI] 4.65
 [HOMOL] SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP (KIAA0108).
 5e-40
 [PROSITE] CK2_PHOSPHO_SITE 3
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 1
 [PROSITE] ASN_GLYCOSYLATION 3
 [KW] SIGNAL_PEPTIDE 49
 [KW] TRANSMEMBRANE 2
 [KW] LOW_COMPLEXITY 20.80 %

SEQ MKMVAPWTRFYSNSCLCCHVRTGILLGVWYLIINAVVLLILLSALADPDQYNFSSSEL
 SEGXXXXXXXXXXXXXXXXXXXXX.....
 PRD ccc
 MEM

SEQ GGDFFEMDDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAI
 SEGXXXXXXXXXXXXXXXXXXXXX.....
 PRD ccc
 MEMMM

SEQ TVLIYPNSIQEYIRQLPPNFPYRDDVMSVNPTCLVLIILLFISIIILTFKGYLISCVWNCY
 SEGXXXXXXXXXXXXXXXXXXXXX.....
 PRD hhhcc
 MEM MMMMM.....MM...

SEQ RYINGRNSSDVLVYVTSNDTTVLLPPYDDATVNGAAKEPPPPYVSA
 SEG
 PRD ecc
 MEM

Prosite for DKFZphutel_24e11.1

PS00001	54->58	ASN_GLYCOSYLATION	PDOC00001
PS00001	187->191	ASN_GLYCOSYLATION	PDOC00001
PS00001	198->202	ASN_GLYCOSYLATION	PDOC00001
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00007	186->195	TYR_PHOSPHO_SITE	PDOC00007

(No Pfam data available for DKFZphutel_24e11.1)

DKFZphute1_24j6

group: cell structure and motility

DKFZphutes1_24j6 encodes a novel 571 amino acid protein with strong similarity to rat cell adhesion regulator (CAR1).

The novel protein is very similar to Car1 and thus seems to be involved in regulation cell-cell adhesion. It contains a RGD cell attachment site.

The new protein can find application in modulation of cell-cell-adhesion.

strong similarity to rat CAR1 A.thaliana T19C21.5

complete cDNA, complete cds, EST hits
potential frame shift at Bp 1241 according to CAR1
but frame shift might be in CAR1 sequence!
ESTs T73366 AA362984 confirm this sequence

Sequenced by Qiagen

Locus: /map="939.9 cR from top of Chr2 linkage group"

Insert length: 3333 bp

Poly A stretch at pos. 3316, no polyadenylation signal found

```
1  ACGCGTCCGA GCTGGCTCAG GCGTCCGCT AGGCTCGGAC GACCTGCTGA
51  GCCTCCCAAA CCGCTTCCAT AAGGCTTTGC CTTTCCAAC TCAGCTACAG
101 TGTTAGCTAA GTTTGGAAAG AAGGAAAAAA GAAAATCCCT GGGCCCTTT
151 TCTTTGTTC TTTGCCAAAG TCGTCGTTGT AGTCTTTTG CCCAAGGCTG
201 TTGTGTTTTT AGAGGTGCTA TCTCCAGTTC CTTGCACTCC TGTAAACAAG
251 CACCTCAGCG AGAGCAGCAG CAGCGATAGC AGCCGCAGAA GAGCCAGCGG
301 GGTGCGCTAG TGTCATGACC AGGGCGGGAG ATCACAACCG CCAGAGAGGA
351 TGCTGTGGAT CCTTGGCCGA CTACCTGACC TCTGCAAAAT TCCTTCTCTA
401 CCTTGGTCAT TCTCTCTCTA CTTGGGGAGA TCGGATGTGG CACTTTGCGG
451 TGCTGTGTTT TCTGGTAGAG CTCTATGGAA ACAGCCTCCT TTTGACAGCA
501 GTCTACGGGC TGGTGGTGGC AGGGTCTGTT CTGGTCCTGG GAGCCATCAT
551 CGGTGACTGG GTGGACAAGA ATGCTAGACT TAAAGTGGCC CAGACCTCGC
601 TGGTGGTACA GAATGTTTCA GTCATCCTGT GTGGAATCAT CCTGATGATG
651 GTTTTCTTAC ATAAACATGA GCTTCTGACC ATGTACCATG GATGGGTTCT
701 CACTTCTGCT TATATCCTGA TCATCACTAT TGCAAAATAT GCAAAATTTG
751 CCAGTACTGC TACTGCAATC ACAATCCAAA GGGATTGGAT TGTGTTGTTT
801 GCAGGAGAAG ACAGAAGCAA ACTAGCAAAT ATGAATGCCA CAATACGAAG
851 GATTGACCAG TTAACCAACA TCTAGCCCC CATGGCTGTT GGCCAGATTA
901 TGACATTGGC CTCCCAGTCT ATCGGCTGTG GCTTTATTTC GGGATGGAAC
951 TTGGTATCCA TGTGCGTGA GTACGTCTCT CTCTGGAAGG TTTACCAGAA
1001 AACCCAGCT CTAGCTGTGA AAGCTGGTCT TAAAGAAGAG GAAACTGAAT
1051 TGAACAGCT GAATTTACAC AAAGATACTG AGCCAAAACC CCTGGAGGGA
1101 ACTCATCTAA TGGGTGTGAA AGACTCTAAC ATCCATGAGC TTGAACATGA
1151 GCAAGAGCCT ACTTGTGCC TCCAGATGGC TGAGCCCTTC CGTACCTTCC
1201 GAGATGGATG GGTCTCCTAC TACAACAGC CTGTGTTTCT GGCTGGCATG
1251 GGTCTTGCTT TCCTTTATAT GACTGTCTTG GGCTTTGACT GCATCACCAC
1301 AGGTACGCC TACACTCAGG GACTGAGTGG TTCCATCCTC AGTATTTTGA
1351 TGGGAGCATC AGCTATAACT GGAATAATGG GAACTGTAGC TTTTACTTGG
1401 CTACGTCGAA AATGTGGTTT GGTTCGGACA GGTCTGATCT CAGGATTGGC
1451 ACAGCTTTCC TGTTTGATCT TGTGTGTGAT CTCTGTATTC ATGCCCTGGAA
1501 GCCCCCTGGA CTTGTCCGTT TCTCCTTTTG AAGATATCCG ATCAAGGTTT
1551 ATCAAGGAG AGTCAATTAC ACCTACCAAG ATACCTGAAA TTACAACCTGA
1601 AATATACATG TCTAATGGGT CTAATCTGCT TAATATTGTC CCGGAGACAA
1651 GTCCTGAATC TGTGCCATA ATCTCTGTCA GTCTGCTGTT TGCAGGCGTC
1701 ATTGCTGCTA GAATCGGTCT TTGGTCCTTT GATTTAACTG TGACACAGTT
1751 GCTGCAAGAA AATGTAATTG AATCTGAAAG AGGCATTATA AATGGTGTAC
1801 AGAATCCCAT GAACTATCTT CTTGATCTTC TGCATTTCAT CATGGTCATC
1851 CTGGCTCCAA ATCCTGAAGC TTTTGGCTTG CTCGTATTGA TTTCACTCTC
1901 CTTTGTGGCA ATGGGCCACA TTATGTATTT CCGATTGGCC CAAAATACTC
1951 TGGGAAACAA GCTCTTTGCT TGCGGTCTCT ATGCAAAAGA AGTTAGGAAG
2001 GAAATCAAG CAAATACATC TGTGTTTGA GACAGTTTAA CTGTTGCTAT
2051 CCTGTACTA GATTATATAG AGCACATGTG CTTATTTTGT ACTGCAGAA
2101 TCCAAATAAT GGCTGGGTGT TTTGCTCTGT TTTTACCACA GCTGTGCCTT
2151 GAGAACTAAA AGCTGTTTAG GAAACCTAAG TCAGCAGAAA TTAATGATT
2201 AATTTCCTT ATGTTGAGGC ATGGAAAAAA AATTGAAAAA GAAAAACTCA
2251 GTTTAAATAC GGAGACTATA ATGATAACAC TGAATCCCC TATTTCTCAT
2301 GAGTAGATAC AATCTTACGT AAAAGAGTGG TTAGTCACGT GAATTCAGTT
2351 ATCATTGAC AGATTCTTAT CTGTACTAGA ATTCAGATAT GTCAGTTTTC
2401 TGCAAACTC ACTCTGTTC AAGACTAGCT AATTTATTTT TTTGCATCTT
2451 AGTTATTTTT AAAACAAAT TCTTCAAGTA TGAAGACTAA ATTTTGATAA
2501 CTAATATTAT CTTATTGAT CCTATTGATC TTAAGGTATT TACATGTATG
```

```

2551 TGGAAAAACA AAACACTTAA CTAGAATTCT CTAATAAGGT TTATGGTTTA
2601 GCTTAAAGAG CACCTTTGTA TTTTATTAT CAGATGGGGC AACATATTGT
2651 ATGAAGCATA GTAGCACTT CACAGCATGG TTATCATGTA AGCTGCAGGT
2701 AGAAGCAAAG CTGTAAGTA GATTTATCAC ACAATGACTG CATAACAGCT
2751 TCAAAATATGT CAATAGTTTG GTCATAGAAC CTAGAAGCCA AAAGCCACAC
2801 AGAAGGGCAA GAATCCCAAT TTAACCTCAT TTATCATCAT TAGTGATCTG
2851 TGTGTAGAA CATGAGGGTG TAAGCCTTCA GCCTGGCAAG TTACATGTAG
2901 AAAGCCACAC CTTGTGAAGG TTTTGTTTTA CAAATCACTT GATTTAACAC
2951 ACTCAGGTAG AATATTTTTA TTTTACTGT TTTATACCCA GAAGTTATTT
3001 CTACATTGTT CTACAGCAAG AATATTCATA AAAGTATCCC TTCAAATGC
3051 CTTTGAGAAG AATAGAAGAA AAAAAGTTTG TATATATTTT AAAAAATTGT
3101 TTTAAAGTC AGTTTGCAAC ATGTCTGTAC CAAGATGGTA CTTTGCCTTA
3151 ACCGTTTATA TGCACCTTCA TGGAGACTGC AATACGTTGC TATGAGCACT
3201 TTCTTTATCC TTGGAGTTTA ATCCTTTGCT TCATCTTTCT ACAGTATGAC
3251 ATAATGATTT GCTATGTTGT AAAATCTTTG TAAAAAATTT CTATATAAAA
3301 ATATTTTGAA AATCTTAAAA AAAAAAAAAA AAA

```

BLAST Results

Entry HS389210 from database EMBL:
human STS SHGC-10164.
Score = 1592, P = 1.5e-64, identities = 346/364

Entry HS933343 from database EMBL:
human STS WI-16551.
Score = 1193, P = 5.7e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 315 bp to 2027 bp; peptide length: 571
Category: strong similarity to known protein

```

1 MTRAGDHNQ RGCCGSLADY LTSAKFLLYL GHSLSTWGDR MWHFAVSVFL
51 VELYNSLLT TAVYGLVVAG SVLVLGAIIG DWVDKNARLK VAQTSLVVQN
101 VSVILCGIIL MMVFLHKHEL LTMVHGWLVT SCYILIITIA NIANLASTAT
151 AITIQRDWIV VVAGEDRSKL ANMNATIRRI DQLTNILAPM AVQIMTFGS
201 PVIGCGFISG WNLVSMCWEY VLLWKVYQKT PALAVKAGLK EEETELKQLN
251 LHKDTEPKPL EGTMLMGVKD SNIHELEHEQ EPTCASQMAE PFRTFRDGVV
301 SYYNQPVFLA GMGLAFLYMT VLGFDCTTIG YAYTQGLSGS ILSILMGASA
351 ITGIMGTVAF TWLRRKCGLV RTGLISGLAQ LSCLILCVIS VFMPGSPDL
401 SVSPFEDIRS RFIQGESITP TKIPEITTEI YMSNGSNSAN IVPETSPESV
451 PIISVSLLEA GVIAARIGLW SFDLTVTQLL QENVIESERG IINGVQNSMN
501 YLLDLLHFM VILAPNPEAF GLLVLISVSF VAMGHIMYFR FAQNTLGNKL
551 FACGPDAKEV RKENQANTSV V

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutell_24j6, frame 3

TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds., N
= 1, Score = 1472, P = 7.2e-151

TREMBL:AC004683_5 gene: "T19C21.5"; Arabidopsis thaliana chromosome II
BAC T19C21 genomic sequence, complete sequence., N = 2, Score = 437, P
= 2.8e-60

TREMBL:AF039046_2 gene: "R09B5.4"; Caenorhabditis elegans cosmid
R09B5., N = 2, Score = 323, P = 1.5e-43

>TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds.
Length = 405

HSPs:

Score = 1472 (220.9 bits), Expect = 7.2e-151, P = 7.2e-151
Identities = 288/319 (90%), Positives = 297/319 (93%)

Query:	1	MTRAGDHNRRQGCCGLADYLTSAKFLLYLGHSLSTWGRDMWHFAVSVFLVELYGNLSLL	60
		MT++ D Q GCCGSLA+YLTSAKFLLYLGHSLSTWGRDMWHFAVSVFLVELYGNLSLL	
Sbjct:	1	MTKSRDQTHQEGCCGLANYLTSAKFLLYLGHSLSTWGRDMWHFAVSVFLVELYGNLSLL	60
Query:	61	TAVYGLVVAGSVLVLGAIGDWDVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL	120
		TAVYGLVVAGSVLVLGAIGDWDVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHK+EL	
Sbjct:	61	TAVYGLVVAGSVLVLGAIGDWDVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKNEL	120
Query:	121	LTMYHGWLVTSCYILIIITIANIANLASTATAITIQRDWIVVVAGEDRSKANMNATIRRI	180
		L MYHGWLVT CYILIIITIANIANLASTATAITIQRDWIVVVAGE+RS+LA+MNATIRRI	
Sbjct:	121	LNMYHGWLVTVCYILIIITIANIANLASTATAITIQRDWIVVVAGENRSRLADMNATIRRI	180
Query:	181	DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK	240
		DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEY LLWKVYQKTPALAVKA LK	
Sbjct:	181	DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYFLLWKVYQKTPALAVKAAKL	240
Query:	241	EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVV	300
		EE+ELKQL KDTEPKPLEGTHLMG KDSNI ELE EQEPTCASQ+AEFPFRTFRDGVV	
Sbjct:	241	VEESELKQLTSPKDTEPKPLEGTHLMGEKDSNIRELECEQEPTCASQTAEPFRTFRDGVV	300
Query:	301	SYYNQPVFLAGMGLAF-LY 318	
		SYYNQPVFL G F LY	
Sbjct:	301	SYYNQPVFLGWHGPGGFPLY 319	

Pedant information for DKFZphutel 24j6, frame 3

Report for DKFZphute1 2416.3

```
[LENGTH]          571
[MW]               62542.72
[pI]               6.08
[HOMOL]            TREMBL:U76714_1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus
norvegicus cell adhesion regulator (CAR1) mRNA, complete cds. 1e-141
[BLOCKS]           BL00341D
[PROSITE]          MYRISTYL      15
[PROSITE]          MITOCH CARRIER 1
[PROSITE]          CK2_PHOSPHO_SITE      6
[PROSITE]          PROKAR_LIPOPROTEIN    1
[PROSITE]          PKC_PHOSPHO_SITE      4
[PROSITE]          ASN_GLYCOSYLATION     4
[PFAM]             Laminin B (Domain IV)
[KW]               TRANSMEMBRANE 4
[KW]               LOW COMPLEXITY      8.76 %
```

```
SEQ      MTRAGDHNRRQGCCGLADYLTSAKFLLYLGHSLSTWGDMMHFAVSFLVELYGNLSLL
SEG      .....
PRD      cccccccccccccccccchhhhhhhheeecccceccccchhhhhhhheeecccceee
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
```

```
SEQ  TAVYGLVVGASVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMVF LHKHEL
SEG  . xxxxxxxxxxxxxxxxxxxx .
PRD  ehhhhhhhhhccccceccccccchhhhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhhhh
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
```

```
SEQ      LTM YHG VLT SCY ILI ITIAN IAN LST ATA ITI QRD WIV VVAGED RSK LAN MN NAT IRRI
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      hhccccchhhhhhhhhhhhhhhhhhhhhheeeccceeeeeeccccchhhhhhhhhhhhhhhhh
MEM      MMMMMMMM
```

```
SEQ      DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK
SEG      .
PRD      hhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhhccccchhhhhhhhh
MEM
```

```
SEQ      EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVWV
SEG      .....
PRD      hhhhhhhhhhhccccccccccccceeeeeeccccccccccccccccccccccccccccce
MEM
```

SEQ SYYNQPVFLAGMGLAFLYMTVLGFDCITTGYAYTQGLSGSILSILMGASAITGIMGTVAF
SEG
PRD eeeccceeeccccchhhhhhccccceeeeeeeccccceeeeeeeccccceeeehhhhhh

```
MEM      .....
SEQ      TWLRRKCGLVRTGLISGLAQLSCLILCVISVFMPPGSLDLSVSPFEDIRSRFIQGESITP
SEG      .....xxx
PRD      hhhhhhccccccccchhhhhhhhhhhhhhhccccccccccccchhhhhcccccccc
MEM      .....
SEQ      TKIPEITTEIYMSNGSNSANIVPETSPEPVPIISVLLFAGVIAARIGLWSFDLTVTQLL
SEG      xxxxxxxxxx.....
PRD      cccccceeeeeeccccccccccccccccceeeeee hhhhhhhhhhhccccchhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....
SEQ      QENVIESERGIINGVQNSMNYLLDLLHFIMVILAPNPEAFGLLVLSVSFVAMGHIMYFR
SEG      .....
PRD      hhhhccccceeeeeeccccchhhhhhhhhheeeeeeccccceeeeeeccccceeee
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....
SEQ      FAQNTLGNKLFACGPDAKEVRKENQANTSVV
SEG      .....
PRD      eccccceeeeeeccccchhhhhhhhhcccccc
MEM      .....
```

Prosites for DKFZphute1_24j6.3

PS000001	100->104	ASN_GLYCOSYLATION	PDOC000001
PS000001	174->178	ASN_GLYCOSYLATION	PDOC000001
PS000001	434->438	ASN_GLYCOSYLATION	PDOC000001
PS000001	567->571	ASN_GLYCOSYLATION	PDOC000001
PS000005	23->26	PKC_PHOSPHO_SITE	PDOC000005
PS000005	176->179	PKC_PHOSPHO_SITE	PDOC000005
PS000005	294->297	PKC_PHOSPHO_SITE	PDOC000005
PS000005	487->490	PKC_PHOSPHO_SITE	PDOC000005
PS000006	16->20	CK2_PHOSPHO_SITE	PDOC000006
PS000006	36->40	CK2_PHOSPHO_SITE	PDOC000006
PS000006	294->298	CK2_PHOSPHO_SITE	PDOC000006
PS000006	396->400	CK2_PHOSPHO_SITE	PDOC000006
PS000006	403->407	CK2_PHOSPHO_SITE	PDOC000006
PS000006	445->449	CK2_PHOSPHO_SITE	PDOC000006
PS000008	12->18	MYRISTYL	PDOC000008
PS000008	65->71	MYRISTYL	PDOC000008
PS000008	76->82	MYRISTYL	PDOC000008
PS000008	193->199	MYRISTYL	PDOC000008
PS000008	267->273	MYRISTYL	PDOC000008
PS000008	311->317	MYRISTYL	PDOC000008
PS000008	336->342	MYRISTYL	PDOC000008
PS000008	339->345	MYRISTYL	PDOC000008
PS000008	353->359	MYRISTYL	PDOC000008
PS000008	368->374	MYRISTYL	PDOC000008
PS000008	373->379	MYRISTYL	PDOC000008
PS000008	435->441	MYRISTYL	PDOC000008
PS000008	461->467	MYRISTYL	PDOC000008
PS000008	490->496	MYRISTYL	PDOC000008
PS000008	494->500	MYRISTYL	PDOC000008
PS000013	122->133	PROKAR_LIPOPROTEIN	PDOC000013
PS000215	404->414	MITOCH_CARRIER	PDOC000189

Pfam for DKFZphute1 24j6.3

HMM_NAME	Laminin B (Domain IV)		
HMM	*YWRIPERFLGDQvTsYGGkLe*		
	Y+R	+ LG+++ + G + +	
Query	538	YFREAO ^{NT} LG ^{NK} LFACG ^{DP} AK	558

DKFZphutel_2h3

group: differentiation/development

DKFZphutel 2h3 encodes a novel 267 amino acid protein, with similarity to ITM2 (integral membrane protein 2) of chicken and mouse.

The novel protein contains a prenyl group binding site (CAAX box) and seems to be post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group. The similar gallus G. protein E25 a marker for chondro-osteogenic differentiation.

The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

strong similarity to mouse E25 and gallus E3-16

complete cDNA, EST hits
complete cds according to E25 start at Bp 56
putative transmembrane protein (1 TM)

Sequenced by AGOWA

Locus: unknown

Insert length: 2033 bp
Poly A stretch at pos. 2007, polyadenylation signal at pos. 1986

```
1 GGACCGAGGC TGCACCGGCA GAGGCTGCGG GCGGACGCG CGGGCCGGCG
51 CAGCCATGGT GAAGATTAGC TTCCAGCCCG CCGTGGCTGG CATCAAGGGC
101 GACAAGGCTG ACAAGGCGTC GCGTCCGGCC CCTGCGCCGG CCTCGGCCAC
151 CGAGATCCTG CTGACGCCGG CTAGGGAGGA GCAGCCCCCA CAACATCGAT
201 CCAAGAGGGG GAGCTCAGTG GCGGCGGTGT GCTACCTGTC GATGGGCATG
251 GTCGTGCTGC TCATGGGCGT CGTGTTCGCC TCTGTCTACA TCTACAGATA
301 CTTCTTTCTT GCACAGCTGG CCCGAGATAA CTTCTCCGCG TGTGGTGTGC
351 TGTATGAGGA CTCCTGTGCC TCCCAGGTCC GGACTCAGAT GGAGCTGGAA
401 GAGGATGTGA AAATCTACCT CGACGAGAAC TACGAGCGCA TCAACGTGCC
451 TGTGCCCCAG TTTGGCGGCG GTGACCCCTG AGACATCATC CATGACTTCC
501 AGCGGGGTCT GACTGCGTAC CATGATATCT CCCTGGACAA GTGCTATGTC
551 ATCGAACTCA ACACCACCAT TGTGCTGCCC CCTCGCAACT TCTGGGAGCT
601 CCTCATGAACT GTGAAGAGGG GGACCTACCT GCCGCAGACG TACATCATCC
651 AGGAGGAGAT GGTGGTCACG GAGCATGTCA GTGACAAAGG GGCCTGGGGG
701 TCCCTCATCT ACCACCTGTG CAACGGGAAA GACACCTACC GGCTCCGGCG
751 CCGGGCAACG CGGAGGCGGA TCAACAAGCG TGGGGCCAAG AACTGCAATG
801 CCATCCGCCA CTTTCGAGAAC ACCTTCGTGG TGGAGACGCT CATCTGCGGG
851 GTGGTGTGAG GCCCTCCTCC CCCAGAACCC CTTGCCGTGT TCCTCTTTTC
901 TCTTTCCAG CTGCTCTCTG GCCCTCCTCC TTCCCTCTGC TTAGCTTGTA
951 CTTTGGACGC GTTCTATAG AGGTGACATG TCTCTCCATT CCTCTCCAAC
1001 CCTGCCCAAC TCCCTGTACC AGAGCTGTGA TCTCTCGGTG GGGGGCCCAT
1051 CTCGTGCTGAC CTGGGTGTGG CGGAGGGAGA GGCATGCTG CAAAGTGTTT
1101 TCTGTGTCCT ACTGTCTTGA AGCTGGGCGT GCCAAAGCCT GGGCCACAG
1151 CTGACCGGCG AGCCCAAGGG GAAGGACCGG TTGGGGGAGC CGGCGCATGT
1201 AGGCCCTGGG CAAGGGGATG GGGCTGTGGG GCGGGGGCGG CATGGGCTTC
1251 AGAAGTATCT GCACAATTAG AAAAGTCTCT AGAAGCTTTT TCTTGGAGGG
1301 TACACTTTCT TCACTGTCCC TATTCTAGA CCTGGGGCTT GAGCTGAGGA
1351 TGGGACGATG TGCCACAGGA GGGACCCACC AGAGCACAAG AGAAGGTGGC
1401 TACCTGGGGG TGTCCAGGG ACTCTGTGAG TGCCTTCAGC CCACCAGCAG
1451 GAGCTTGGAG TTTGGGGAGT GGGGATGAGT CCGTCAAGCA CAACTGTTCT
1501 CTGAGTGGAA CCAAAGAAGC AAGGAGCTAG GACCCCAAGT CCTGCCCCC
1551 AGGAGCACAA GCAGGGTCCC CTCAGTCAAG GCAGTGGGAT GGGCGGCTGA
1601 GGAACGGGGC AGGCAAGGTC ACTGCTCAGT CACGTCCACG GGGGACGAGC
1651 CGTGGGTCTT GCTGAGTAGG TGGAGCTCAT TGCTTCTCC AAGCTTGGAA
1701 CTGTTTTGAA AGATAACACA GAGGGAAAGG GAGAGCCACC TGGTACTTGT
1751 CCACCCTGCC TCCTCTGTTC TGAAATTCCA TCCCCTCAG CTTAGGGGAA
1801 TGCACCTTTT TCCCTTTCTT TCTCACTTTT GCATGTTTTT ACTGATCATT
1851 CGATATGCTA ACCGTTCTCA GCCCTGAGCC TTGGAGAGGA GGGGTGTAAC
1901 GCCTTCAGTC AGTCTCTGGG GATGAACTC TTAATGCTT TGTATATTTT
1951 CTCATTAGA TCTCTTTTCA GAAGTGTCTA TAGAACATA AAAATCTTTT
2001 ACTTCTGAAA AAAAAAAAAA AAAAGGGCGG CCG
```

BLAST Results

Entry B64417 from database EMBL:
CIT-HSP-2023A7.TR CIT-HSP Homo sapiens genomic clone 2023A7.
Length = 715
Plus Strand HSPs:

Score = 1546 (232.0 bits), Expect = 7.8e-64, P = 7.8e-64
Identities = 310/311 (99%)

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation using cDNA library subtraction.
Molecular cloning and characterization of a gene belonging to a novel multigene family of integral membrane proteins.

Peptide information for frame 2

ORF from 56 bp to 856 bp; peptide length: 267
Category: strong similarity to known protein

```

1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
51 RGSSVGGVCY LSMGMVVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VPVPQFGGGD PADIIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRNEF ELLMNVKRGY YLPQTYIIQE
201 EMVVEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRRIN KRGAKNCNAI
251 RHFENTFVVE TLICGVV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_2h3, frame 2

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16)., N = 1, Score = 573, P = 1.3e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N = 1, Score = 560, P = 3.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1, Score = 456, P = 3.3e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
Length = 262

HSPs:

Score = 573 (86.0 bits), Expect = 1.3e-55, P = 1.3e-55
Identities = 117/264 (44%), Positives = 172/264 (65%)

```

Query: 1 MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY 60
      MVK+SF A+A + A+K ++ ++L+ P ++P G
Sbjct: 1 MVKVSFNSALA--HKEAANKEEENS-----QVLILPPDAKEPEDVVVPAGHKRAWCWC 51

Query: 61 LSMGMVVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSLSS----SQVRTQM-- 112
      + G+ +L G++ Y+Y+YF Q + CG+ Y ED LS +Q+++
Sbjct: 52 MCFGLAFMLAGVILGGAYLYKYFAFQQ---GGVYFCGIKYIEDGLSLPESGAQLKSARYH 108

Query: 113 ELEEDVKIYLDENYERINVPVPQFGGGDPADIIHDFQRGLTAYHDISLDKCYVIELNTTI 172
      +E++++I +E+ E I+VPVP+F DPADI+HDF R LTAY D+SLDKCYVI LNT++
Sbjct: 109 TIEQNIQILEEEDVEFISVPVPEFADSDPADIVHDFHRRLTAYLDLSLDKCYVIPLNTSV 168

Query: 173 VLPFRNFWEELLMNVKRGTYLPQTYIIQEEMVVEHVSDKEALGSFIYHLCNGKDTYRLRR 232
      V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+R
Sbjct: 169 VMPPKNFLELLINIKAGTYLPQSYLIHEQMIVTDRIENVQDLGFFIYRLCRGKETYKLQR 228

Query: 233 RATRRRRINKRGAKNCNAIRHFENTFVETLIC 264
      + + I KR A NC IRHFEN F +ETLIC
Sbjct: 229 KEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260

```

Pedant information for DKFZphut1_2h3, frame 2

[illegible]

PS000001	169->173	ASN_GLYCOSYLATION	PDOC000001
PS000004	50->54	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	187->191	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	232->236	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	49->52	PKC_PHOSPHO_SITE	PDOC000005
PS000005	209->212	PKC_PHOSPHO_SITE	PDOC000005
PS000005	227->230	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000006	30->34	CK2_PHOSPHO_SITE	PDOC000006
PS000006	110->114	CK2_PHOSPHO_SITE	PDOC000006
PS000006	209->213	CK2_PHOSPHO_SITE	PDOC000006
PS000007	119->127	TYR_PHOSPHO_SITE	PDOC000007
PS000008	52->58	MYRISTYL	PDOC000008
PS000008	71->77	MYRISTYL	PDOC000008
PS000008	138->144	MYRISTYL	PDOC000008
PS000008	243->249	MYRISTYL	PDOC000008
PS00294	264->268	PRENYLATION	PDOC000266

542

DKFZphmcfl_1a11

group: transmembrane protein

DKFZphmcfl_1a11 encodes a novel 393 amino acid protein with weak similarity to S.pombe SPBC29A3_3 protein and S. cerevisiae putative membrane protein YDR255c.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes and as a new marker for mammary carcinoma cells.

similarity to YDR255c and SPBC29A3.03c

membrane regions: 1

Summary DKFZphmcfl_1a11 encodes a novel 393 amino acid protein, with similarity to YDR255c and SPBC29A3.03c.

similarity to YDR255c and SPBC29A3.03c

complete cDNA, complete cds, EST hits
potential start at Bp 110 matches kozak consensus

Sequenced by DKFZ

Locus: /map="542.7 cR from top of Chr5 linkage group"

Insert length: 1819 bp

Poly A stretch at pos. 1808, no polyadenylation signal found

```

1  CCCGGCCCCAG  CCCCCGAAGA  GCCGCCTCAG  CCGGGGGGAG  TTGCTCGGAC
51  TCAAACGTCCT  AGTCCTCGTG  CGACCGCGCT  GGGTCGGAAG  TGAGCAGGCT
101  CAGGCCACCA  TGGAGCAGTG  TGCGTGCGTG  GAGAGAGAGC  TGGACAAGGT
151  CCTGCAGAAG  TTCCTGACCT  ACGGGCAGCA  CTGTGAGCGG  AGCCTGGAGG
201  AGCTGCTGCA  CTACGTGGGC  CAGCTGCGGG  CTGAGCTGGC  CAGCGCAGCC
251  CTCCAGGGGA  CCCCTCTCTC  AGCCACCCTC  TCTCTGGTGA  TGTACAGTGT
301  CTGCCGGAAG  ATCAAAGATA  CGGTGCAGAA  ACTGGGCTCG  GACCATAAGG
351  ACATTCACAG  CAGTGTATCC  CGAGTGGGCA  AAGCCATTGA  CAGGAACCTC
401  GACTCTGAGA  TCTGTGGTGT  TGTGTACAGT  GCGGTGTGGG  ACGCGCGGGA
451  ACAGCAGCAG  CAGATCCTGC  AGATGGCCAT  CGTGAACAC  CTGTATCAGC
501  AGGGCATGCT  CAGCGTGGCC  GAGGAGCTGT  GCCAGGAATC  AACGCTGAAT
551  GTGGACTTGG  ATTTCAAGCA  GCCTTTCCTA  GAGTTGAATC  GAATCCTGGA
601  AGCCTGACAC  GAACAAGACC  TGGGTCCCTG  GTTGAATGG  GCCGTCTCCC
651  ACAGGCAGCG  CCTGCTGGAA  CTCACAGCT  CCCTGGAGTT  CAAGCTGCAC
701  CGACTGCACT  TCATCCGCCT  CTTGGCAGGA  GGCCCCGCGA  AGCAGCTGGA
751  GGCCCTCAGC  TATGCTCGGC  ACTTCCAGCC  CTTTGCTCGG  CTGCACCAGC
801  GGGAGATCCA  GGTGATGATG  GGCAGCCTGG  TGTACCTGCG  GCTGGGCTTG
851  GAGAACTCAC  CCTACTGCCA  CCTGCTGGAC  AGCAGCCACT  GGGCAGAGAT
901  CTGTGAGACC  TTTACCCGGG  ACGCCTGTTC  CCTGCTGGGG  CTTTCTGTGG
951  AGTCCCCCCT  TAGCGTCAGC  TTTGCCTCTG  GCTGTGTGGC  GCTGCCTGTG
1001  TTGATGAACA  TCAAGGCTGT  GATTGAGCAG  CGGCAGTGCA  CTGGGGTCTG
1051  GAATCACAAAG  GACGAGTTAC  CGATTGAGAT  TGAAGTAGGC  ATGAAGTGCT
1101  GGTACCACTC  CGTGTTCGCT  TGCCCCATCC  TCCGCCAGCA  GACGTCAGAT
1151  TCCAAACCCTC  CCATCAAGCT  CATCTGTGGC  CATGTTATCT  CCCGAGATGC
1201  ACTCAATAAG  CTCATTAATG  GAGGAAAGCT  GAAAGTGCCC  TACTGTCCCA
1251  TGGAGCAGAA  CCCGGCAGAT  GGGAAACGCA  TCATATTCG  ATTCTACCT
1301  GGAAGGAATT  TTGTGAAAG  GGGTTTTCAC  CTGTGAGCCT  TGGTCTGTCT
1351  CGGTAGGGTG  GTCAACTTCA  GTGGACTGTG  GTTGGTTTCA  GAGCGCCTGG
1401  CTGAGGAGTT  CCACTGAGGG  GAGCACTGGA  GCAGCCCTTT  GGCAGAGGCT
1451  GAGGAGGGAG  ATGGACCAGC  CCACGCCTGG  CACCTGGCTC  CATGGCATAA
1501  GGAAGGGAG  ATGCTGGCCT  CTGTGCTCCT  GCTGTCTTTT  CCTGTTTCTG
1551  TTTGCGTTTG  ACTTAGTAGC  AACCAGACAG  GTGGCAAGGG  ATTTGGTCTT
1601  CAGCAGTAGA  CATCTTCCA  CCCCTGCCCT  CAGCCAAGTC  TCTTGCTGCC
1651  ATGCCAATGC  TATGTCCACC  CTTGCCCTC  GGCCCAAGAG  TGTCACGCGG
1701  TGGCCACCT  CTTCTCCCA  CTACAGCCTC  AACAGTATGT  ACCATCTCCC
1751  ACTGTAAATA  GTCCCAAGTA  GAACGGAATG  CCGTTGTTTT  ATAACTTTGA
1801  ACAATGTAA  AAAAAAAA

```

BLAST Results

Entry HS579359 from database EMBL:
human STS WI-6350.

Score = 1027, P = 9.9e-40, identities = 207/209

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 1288 bp; peptide length: 393
 Category: similarity to unknown protein

```

1 MEQCACVERE LDKVLQKFLT YGQHCERSLE ELLHYVGQLR AELASAAALQG
51 TPLSATLSLV MSQCCRKIKD TVQKLASDHK DIHSSVSRVG KAIDRNFDSE
101 ICGVVSDAVW DAREQQQIIL QMAIVEHLYQ QGMLSVAEEL CQESTLNVDL
151 DFKQPFLELN RILEALHEQD LGPALEWAVS HRQRLELNS SLEFKLHRLH
201 FIRLLAGGPA KQLEALSYAR HFQPFARLHQ REIQVMMGSL VYLRGLGLEKS
251 PYCHLLDSSH WAEICETETR DACSLGLSV ESPLSVSFAS GCVALPVLMM
301 IKAVIEQRQC TGVWNHKKDEL PIEIELGMKC WYHSVFACPI LRQQTSDSNP
351 PIKLICGHVI SRDALNKLIN GGKLKCPYCP MEQNPADGKR IIF

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_lall, frame 2

TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3., N = 2, Score = 302, P = 3.4e-42

PIR:S67312 probable membrane protein YDR255c - yeast (Saccharomyces cerevisiae), N = 1, Score = 271, P = 5.3e-22

TREMBL:CET07D1_2 gene: "T07D1.2"; Caenorhabditis elegans cosmid T07D1., N = 1, Score = 193, P = 5.6e-13

>TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3.
 Length = 398

HSPs:

Score = 302 (45.3 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 55/142 (38%), Positives = 89/142 (62%)

Query: 252 YCHLLDSSHAEICETFTRDACSLGLSVESPLSVSFASGCVALPVLMMNIKAVIEQRQCT 311
 Y +LD W + F R+ C+ LG+S+ESPL + +G +ALP+L+ ++++++

Sbjct: 258 YIDVLDLD-WKSLELLFVREFCAALGMSLESPLDIVNAGAIAPILLKMSSIMKKKHTE 316

Query: 312 GVVWNHKKDELPIEIELGMKCWYHSVFACPI LRQQTSDSNPPIKLICGHVISRDALNKLING 371
 W + ELP+EI L +HSVF CP+ ++Q ++ NPP+ + CGHVI +++L +L

Sbjct: 317 --WTSQGELPVEIFLPSSYHFHSVFTCPVSKEQATEENPPMMSCGHVIVKESLRQLSRN 374

Query: 372 G--KLKCPYCPMEQNPADGKRIIF 393
 G + KCPYCP E AD R+ F

Sbjct: 375 GSQRFKCPYCPNENVAADAIRVYF 398

Score = 161 (24.2 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 51/221 (23%), Positives = 102/221 (46%)

Query: 22 GQHCERSLEELLHYVGQLRAELASAAALQGTPLSATLSLVMSQCCRKIKDTVQKLASDHKD 81
 G C L EL + + + L+ P ++ LV C K + L K

Sbjct: 15 GNKCLAKLNEL---ESILKDAKSKLKD-PTTSMKELVA--CSEKTQQVFDDLRKTEK 67

Query: 82 IHSSVSRVGKAIDRNFDSEICGVVSDAVWDAREQQQIILQMAIVEHLYQQGMLSVAEELC 141
 H+S++R GK +++ F+ ++ + +++++++ + A+ H ++QG + +A C

Sbjct: 68 FHTSLNRFGKTLKFNFDLEDIKLHSSFESKKRE---IDTALS LHFRRQGDVELAHLFC 124

Query: 142 QESTLNVDLDFKQPFLELNRIEALHEQDLGPALEWAVSHRQRLELNS SLEFKLHRLHF 201
 +E+ + + F L I++ + ++DL +EWA R L SSLE+ L +

Sbjct: 125 KEAGIEEPSSESLHVFTLLKSIVQGIRDKDLKPIEWASQCRGYLERKGS SLEYTLQKYRL 184

Query: 202 IRLLAGGPAQL-EALSYAR-HFQPFARLHQREIQVMMGSLVY 242
 + K + A+ Y R + F + H +IQ M +L +

Report for DKFZphmcf1 1a11.2

[illegible]

PS000001	189->193	ASN_GLYCOSYLATION	PDOC000001
PS000005	180->183	PKC_PHOSPHO_SITE	PDOC000005
PS000006	28->32	CK2_PHOSPHO_SITE	PDOC000006
PS000006	135->139	CK2_PHOSPHO_SITE	PDOC000006
PS000006	190->194	CK2_PHOSPHO_SITE	PDOC000006
PS000007	211->219	TYR_PHOSPHO_SITE	PDOC000007
PS000007	27->36	TYR_PHOSPHO_SITE	PDOC000007
PS000007	244->253	TYR_PHOSPHO_SITE	PDOC000007
PS000008	37->43	MYRISTYL	PDOC000008
PS000008	50->56	MYRISTYL	PDOC000008
PS000009	387->391	AMIDATION	PDOC000009
PS000013	282->293	PROKAR LIPOPROTEIN	PDOC000013

545

DKFZphmcf1_1c23

group: mammary carcinoma derived

DKFZphmcf1_1c23.1 encodes a novel 311 amino acid proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of mamma carcinoma-specific genes.

unknown, proline rich protein

complete cDNA, complete cds? potential start at Bp 50, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3077 bp

Poly A stretch at pos. 3067, polyadenylation signal at pos. 3048

```
1  AACTGGCCCC CTCCCCACC CCCTGCCCTT GAGGAGCAGG ACCTGTCCAT
51  GGCTGACTTC CCCCCACCAG AGGAGGCTTT TTTCTCTGTG GCCAGCCCTG
101 AGCCTGCAGG CCCTTCAGGC TCCCCAGAGC TTGTCTAGTC CCCGGCTGCT
151 TCGTCTCTCT CAGCTACTGC TTTGCAGATT CAGCCCCCGG GTAGCCCAAG
201 CCCTCTCTCA GCTCCGCCAG CCCCAGCTCC TGCTAGTTCC GCCCCAGGGC
251 ATGTGGCCAA GCTCCCTCAG AAGGAACCGG TGGGCTGTAG CAAGGGTGGT
301 GGGCTTCCCA GGGAGGACGT AGGTGCGCCC CTGGTCACGC CCTCGCTCCT
351 GCAGATGGTG CGGCTGCGCT CCGTGGGTGC TCCAGGAGGG GCTCCCACCC
401 CAGCACTGGG GCCATCGGCC CCCCAGAAAC CACTGCGAAG GGCCCTGTCA
451 GGGCGGGGCA GCCCAGTGCC TGCCCCCTCC TCAGGGCTCC ATGCTGCGGT
501 CCGACTCAAG GCCTGCAGCC TGGCCGCCAG TGAAGGCCTC TCAAGTGCTC
551 AGCCCAACGG ACCGCCTGAG GCAGAGCCAC GGCCCTCCCA GTCCCTTGCC
601 TCAACGGGCA GTTTCATCTT CTCCAAGGGC TCTAGGAAGC TGCAGCTGGA
651 GCGGCCCCTG TCCCTTGAGA CCCAGGCTGA CCTCCAGCGG AATCTGGTGG
701 CAGAACTCCG GAGCATCTCA GAGCAGCGGC CACCCAGGGC CCAAAGAAG
751 TCACCTAAGG CTCCCCCACC TGTGGCCCGC AAGCCGTCTG TGGGAGTCCC
801 CCCACCCGCC TCCCCCAGTT ACCCTCGAGC TGAGCCCTTT ACTGCTCCTC
851 CCACCAATGG GCTCCCTCAC ACCCAGGACA GGAATAAGAG GGAGCTGGCG
901 GAGAAATGGG GTGTCTTGCA GCTGGTGGGC CCAGAGGAGA AGATGGGCCT
951 CCGGGGCTCA GACTCACAGA AAGAGCTGGC CTGACCACCA GGCACCTCAC
1001 TGGGACTGCT GACCCATCCC AGAAACACAA TCTCAGGGAC CCGAGCAGCT
1051 CCAAGGACGA GAGGATACAG CAGACACAAC CTAATAGAGA GGGCGCCTGC
1101 AGCCTTAACC TCCACGGCCT TCGATACTTA TGCAAGCCTG GTGTTGCTCC
1151 TGTCTCTAGA GTCATCTGCT GCTCATGCCT TTTCCCGAAT GGGTTACACT
1201 CTGGCAGTTG CCGCTTCAGT CTTGGCCTTA GCCTCATCTT GAAAGTGGGA
1251 GCTGGCGGGA GAGGGTGGCT GCGCCCCCTG CTGGCCCTGA GGCTGCAGAG
1301 TTGGGAGAGG GACACCTCAC CTGAGTTTCA TTTTTTTTCA TGTCCAAACC
1351 ATGCACATAT TATAGTCCAG AATCAAAACA CTTTTGAAAA GTGGCTGCAT
1401 GGCCATCTCT CAGGGCCAGG GAAAGTTGAT TCCAAGGGCC TGTTTACATG
1451 GCAGCAGAAAT CCATCCCCGG CAGTCAGCCC ATAGCTTGGG ACCAGTCTGT
1501 GGCCTTCCTG CAGTCCAGT TACTCTCTCT TGGTTCCTGA AGGTGGCCAA
1551 GTCATTGTGT TCCCACAGGC TTCTCTAGGC TGGGGGCAGG TGTGGGGCTG
1601 TGGAAATCCA AAGCACAAAA GGTGCAGAGG GGATTGGCCT TCCTGTGCCT
1651 CAACTCACCA ACCACCTTCC TGCCTTCCAG TTCTGCCAGG TGCTCCATGC
1701 TGGGGACAAG TAGGAGACTG CCAGGGCCCA AAGAAATGGG TGAGCAGTAG
1751 AGTCATCTCG GGGCACTTGG CAGTGTCAGG CACCTGCCCC TGCTCTCCTT
1801 GACCACACTG GGGTGGGTGG GCCCCAGCA CTTAGAGGCG AGGAGCCTTT
1851 TGGCTGAGCA AGCACTGAGG AGGTGGATGG AAGGGAGCAT CTGGAGGGGG
1901 GGAGCTTCTT TGAGCAGTGG GCCCAGGCCT GGCCCTCCAC ACTTCATTCT
1951 CTGACCTTTC TCTCTCTTCA TTTCCGGTGA TGCTCTTCTT GCAGCTGCCT
2001 TTCAGCACAG GTGGTTCCAC TGGGGGCAGC TAACGCTGAG TGACAAGGAT
2051 GGGAGGCCAC AGGTGCATTT TACTCAAGTC TTCTCTAGTC AATGAGGGGC
2101 ACCCAGTGCT TCTAGGGCAG GCTGGGTGGT GGTCCCTTAG GTATCAGCCT
2151 CTCTTACTGT ACTCTCCGGG AATGTTAACC TTTCTATTTT CAGCCTGTGC
2201 CACCTGTCTA GGCAGGCTGG CTTCCCCATT GGCCCTGTGT GGTCCACAGC
2251 AGCGTGGCTG CCCCCAGGG CCACCGCTTC TTTCTTGATC CTCTTCTCTT
2301 AACAGTGACT TGGGCTTGAG TCTGGCAAGG AACCTTGCTT TTAGCTTCAC
2351 CACCAAGGAG AGAGGTTGAC ATGACCTCCC CGCCCCCTCA CCAAGGCTGG
2401 GAACAGAGGG GATGTGGTGA GAGCCAGGTT CCTCTGGCCC TCTCCAGGGT
2451 GTTTTCCACT AGTCACTACT GTCTTCTCCT TGTAGCTAAT CAATCAATAT
2501 TCTTCCCTTG CCTGTGGGCA GTGGAGAGTG CTGCTGGGTG TACGCTGCAC
2551 CTGCCCCACTG AGTTGGGGAA AGAGGATAAT CAGTGAGCAC TGTCTGTCTC
2601 AGAGCTCCTG ATCTACCCCA CCCCCTAGGA TCCAGGACTG GGTCAAAGCT
2651 GCATGAAACC AGGCCCTGGC AGCAACCTGG GAATGGCTGG AGGTGGGAGA
2701 GAACCTGACT TCTCTTTCCC TCTCCCTCCT CCAACATTAC TGGAACTCTA
```

```

2751 TCCTGTTAGG ATCTTCTGAG CTTGTTTCCC TGCTGGGTGG GACAGAGGAC
2801 AAAGGAGAAG GGAGGGTCTA GAAGAGGCAG CCCTTCTTGG TCCTCTGGGG
2851 TAAATGAGCT TGACCTAGAG TAAATGGAGA GACCAAAAGC CTCTGATTTT
2901 TAATTTCAT AAAATGTTAG AAGTATATAT ATACATATAT ATATTCTTTT
2951 AAATTTTGA GTCTTTGATA TGTCTAAAAA TCCATTCCCT CTGCCCTGAA
3001 GCCTGAGTGA GACACATGAA GAAAACTGTG TTTCATTATA AGATGTTAAT
3051 TAAATGATTG AAACCTGAAA AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 49 bp to 981 bp; peptide length: 311
 Category: putative protein
 Classification: unset

```

1 MADFFPPEEA FFSVASPEPA GPSGSPPELV SPAASSSSAT ALQIQPPGSP
51 DPPPAPPAPA PASSAPGHVA KLPQKEPVGC SKGGGPPRED VGAPLVTPSL
101 LQMVLRSVG APGGAPTAL GPSAPQKPLR RALSGRASPV PAPSSGLHAA
151 VRLKACSLAA SEGLSSAQPN GPPEAEPRPP QSPASTASFI FSKGSRKLQL
201 ERPVSPETQA DLQRNLVAEL RSISEQRPQ APKSPKAPP PVARKPSVGV
251 PPPASPSYPR AEPLTAPPTN GLPHTQDRTK RELAENGGLV QLVGPEEKMG
301 LPGSDSQKEL A

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1c23, frame 1

PIR:S49915 extensin-like protein - maize, N = 1, Score = 215, P = 6.1e-15

PIR:A28996 proline-rich protein M14 precursor - mouse, N = 1, Score = 191, P = 3.8e-13

>PIR:S49915 extensin-like protein - maize
 Length = 1,188

HSPs:

Score = 215 (32.3 bits), Expect = 6.1e-15, P = 6.1e-15
 Identities = 81/269 (30%), Positives = 115/269 (42%)

```

Query: 5 PPPEEAFFS----VASPEPAGPSGSPPELVSSPAASSSSATALQIQPPGSP--DPPP---A 55
      PPP S V SP P P SP PA +SS ++ PP +P PPP +
Sbjct: 598 PPPPAPVASPPPPVKSPPPPTPVASPP--PPAPVASSPPPMKSPPPPTPVSSPPPPPEKS 654

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGC SKGGGPPREDVGAPLVTPSL LQMVLRSVGAPGGA 115
      PP P PA S P + P P K PP + + P + PS + P
Sbjct: 655 PPPPPAKSTPPP-EEYPT--PPTSVKSSPPPEKSLPPPTLIPSPPPQEKTPPSTPSKP 711

Query: 116 PTPALGPSAPQKPLRRA-LSGRASVPVAPSSGLHAAVRLKACSLAA SEGLSSAQPN GPPE 174
      P+ PS P++P+ + ++SP PAP S +LA S + + PP
Sbjct: 712 PSSPEKPSPPKEPVSSPPQTPKSSPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPA 771

Query: 175 AEPRPPQSPASTASFIFSKGSRKLQLERPV-SPETQADLQRNLVAELRSISEQRPQAPK 233
      PP +P +S +Q+ P +P++ L V+ + + PP AP
Sbjct: 772 PLSSPPAPQVKSS-----PPPVQVSSPPAPKGSPLAP--VSSPPQVEKTSPPPPAPL 823

Query: 234 KSPKAPPPVARKPSVGV-PPPASPSYPRAEPLTAPPTNGLP 273
      SP P + P V V PPP S P P+++PP P
Sbjct: 824 SSPPLAPK-SSPPHVVSPPPPVVKSSPPAPVSSPPLTPKP 864

Score = 206 (30.9 bits), Expect = 9.1e-14, P = 9.1e-14

```

Identities = 82/261 (31%), Positives = 108/261 (41%)

Query: 17 PEPAG-PSGSPVLVSSPAASS---SSATALQIQPGSPDPPFPAP---PAPAPASSAPGHV 69
 P P G P SP + PAAS+ S T + P P+P P P P P +P
 Sbjct: 410 PTPGGGPPSSP-VPGKPAASAMPSPHTPPDVSPEPLPEPSVPVAPAPMPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTALGPSAPQKP 128
 +P PV G S P V P + +V+L AP G+P P + ++P P
 Sbjct: 469 DYVPPTPPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTSPPAP 528

Query: 129 LRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSPASTAS 188
 + G SP P P S + +K+ A G + P PPE P PP AS
 Sbjct: 529 I-----GSPSP-PPVSVVSPPPVKSPPPPAPVG---SPP--PPEKSPPPPAPVASPPP 577

Query: 189 FIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPVARKPS- 247
 + S L P P ++ VA + PP P SP P PVA P
 Sbjct: 578 PVKSPPPPPTLVASPP--PPVKSPPPPAPVASPPPPVKSPPPPPTVASPPPPAPVASSPPP 635

Query: 248 VGVPPP---ASPSYPRAEPLTAPPTNGLPHTQD 277
 + PPP +SP P P PP P ++
 Sbjct: 636 MKSPPPPPTPVSSPPPEKSPPPPPPAKSTPPPEE 669

Score = 202 (30.3 bits), Expect = 2.9e-13, P = 2.9e-13

Identities = 81/254 (31%), Positives = 110/254 (43%)

Query: 16 SPEPAGPSGSPVLV---SSP--AASSSSATALQIQPGSP-DPPFPAPAPAPASSAPGHVA 70
 SP PA P SP L SSP SS ++ PP +P PP P PA S P HV+
 Sbjct: 817 SPPPA-PLSSPPLAPKSSPPHVVSPPPPVKSPPPPAPVSSPPLTPKPA---SPPAHVS 872

Query: 71 KLPQ---KEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTALGPSAPQ 126
 P+ P + PP E +P TP L ++S P +P + P +
 Sbjct: 873 SPPEVVKPSTPPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPPPAMVSSPMTPKSSP 932

Query: 127 KPLRRAL---SGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSP 183
 P+ + + ++SP PAP S A K+ A L P PPE + PP +P
 Sbjct: 933 PPVVSSPPPTVKSSPPAPVSSPPAPT--KSSPPAPVNL---P--PPEVKSSPPPTP 984

Query: 184 ASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPVA 243
 S+ + P PE ++ V+ + PP AP SP PPPV
 Sbjct: 985 VSSPPAPKSSPPPPAPMSSPPPEVKSSPPPPAPVSSPPPPVKSPPPPAPVSSP--PPPVK 1042

Query: 244 RKPS---VGVPPPASPSYPRAEPLTAPP 268
 P V PPP S P P+++PP
 Sbjct: 1043 SPPPPAPVSSPPPPVKSPPPPAPISSPP 1070

Score = 190 (28.5 bits), Expect = 7.9e-12, P = 7.9e-12

Identities = 74/264 (28%), Positives = 111/264 (42%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLVSSPAAS---SSATALQIQPGSPDPPFPAPAPAS 63
 PPP S PE + P P + P + T+++ PP PP P+P
 Sbjct: 639 PPPPTPVSSPPPEKSPPPPPPAKSTPPPEEYTPPTSVKSSPPPEKSLPPPTLIPSPPP 698

Query: 64 SAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTALGPS 123
 P K P K PP+E V +P TP V +P PTP P
 Sbjct: 699 QEKPTPSTPSKPPSSPEKPS-PPKEPVSSPPQTPK--SSPPAPVSSP--PPTPVSSPP 753

Query: 124 APQKPLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSP 183
 A P+ S ++SP PAP S A ++K+ + + + P PP + PP +P
 Sbjct: 754 A-LAPVSSPPSVKSSPPAPLSSPPAPQVKS----SPPPVQVSSP--PPAPKSSPPLAP 806

Query: 184 ASTASFIFSKGSRKLQLERP-VSPETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPV 242
 S+ + L P ++P++ +V+ + + PP AP SP P
 Sbjct: 807 VSSPPQVEKTSPPAPLSSPPLAPKSSPP--HVVVSSPPPVKSSPPPPAPVSSPPLTPK 864

Query: 243 ARKPS-VGVPP---PASPSYPR-----AEPLTAPP 268
 A P+ V PP P++P P +EP ++PP
 Sbjct: 865 ASPPAHVSSPPEVVKPSTPPAPTTVISPPSEPKSSPP 901

Score = 189 (28.4 bits), Expect = 1.0e-11, P = 1.0e-11

Identities = 86/271 (31%), Positives = 112/271 (41%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLVSSPAAS---SSATALQIQPGSPDPPFPAPAPAS 56
 PPP A S P P S P + VSSP A SS A PP PPPAP
 Sbjct: 768 PPP--APLSSPPAPQVKSPPPVQVSSPPAPKSSPPLAPVSSPPQVEKTSPPPPAPLSS 825

Query: 57 PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAP 116
 P AP SS P V P PV S PP V +P +TP V +P
 Sbjct: 826 PPLAPKSSPPHVVSPP--PVKSS---PPAPVSSPPLTPKAPASPPA--HVSSPPEVV 878

Query: 117 TPALGPSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKAC-SLAASEGL---SSAQP--- 169
 P+ P AP + ++SP P P S V+ ++ +S + SS P

Sbjct: 879 KPST-PPAPTTVISPPSEPKSSPPPTFVSLPPPIVKSSPPFAMVSSPPMTPKSSPPV 937

Query: 170 -NGPPEAEPRPPQSPASTASFIFSKGRKQLERPVSPETQADLQRNLVAELRSISEQR 228
+ PP + PP + P S + + P PE ++ V+ + P

Sbjct: 938 SSPPTVKSSPPAPVSSPPATPKSSPPAPVNL-PPEVKSSPPPTFVSSPPAPKSSP 996

Query: 229 PQAPKKSPPKAPPPVARKPS----VGVPASPSPYRAEPLTAPP 268
P AP SP PPP + P V PPP S P P+++PP

Sbjct: 997 PPAPMSSP--PPPEVKSSPPAPVSSPPPVKSSPPAPVSSPP 1038

Score = 181 (27.2 bits), Expect = 8.8e-11, P = 8.8e-11
Identities = 73/277 (26%), Positives = 105/277 (37%)

Query: 3 DFPPEEAEFFSVASPEPAGPSGSELVSSPAASSSSATALQIQPF---GSPDPP---PA 55
D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +

Sbjct: 469 DYVPPTTP---VPGKSPATSPSPQ-VQPPAASPPPSLVKLSPPQAPVGSPPPVKTTS 524

Query: 56 PPAPAPASSAPGHVAKL----PQKEPVGCSKGGPPREDVGAPLVTPSLQMVRLRSVGA 111
PPAP + S P V+ + P K P + G PP + P P ++S

Sbjct: 525 PPAPIGSPSPPPVSVVSPPPPVKSSPPAPVGSPPPEKSPPPAPVASPPPVKSSPP 584

Query: 112 PG--GAPTALGPSAPQKPLRRA---LSGRASVPVAPSSGLHAAVRLKACSLAASEGLSS 166
P +P P + P P+ + P P S A V + + +

Sbjct: 585 PTLVASPPPVKSSPPAPVASPPPVKSSPPPTPVASPPPPAPVASSPPPMKSSPPPTP 644

Query: 167 AQPNGPPEAEPRPPQSPASTASFIFSKGRKQLERPVSPETQADLQRNLVAELRSISEQ 226
PPE P PP PA + + ++ PE L+ +

Sbjct: 645 VSSPPPEKSP-PPPPAKSTPPPEEYPTPTSVKSSPPPEKSLP-PPTLIPSPPPQEK 702

Query: 227 RPPQAPKKSPPKAPP-PVARKPSVGVPASPSPYRAEPLTAPP 268
PP P K P +P P K V PP S P P+++PP

Sbjct: 703 TPTSTPSKPPSSPEKSPPKPVSSPPQTPKSSPPAPVSSPP 745

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 78/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAEFFSVASPEPAGP---SGSELVSSPAASSSSATALQIQPGSP--DPPAP-- 56
PPP +P+PA P S PE+V P+ + T I PP P PPP P

Sbjct: 850 PPAAPVSSPPLTKPASPPAHVSSPPEVK-PSTPPAPTTV--ISPPSEPKSSPPPTVS 906

Query: 57 -PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLQMVRLRSVGA 115
P P SS P + P P PP V +P P++ V +P

Sbjct: 907 LPPPIVKSSPPFAMVSSPPMTPKS-----SPPPVVSSP--PPTVKSSPPAPVSSPPAT 959

Query: 116 PTPALGPSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEA 175
P + P+ P ++SP P P S A + S +SS P PPE

Sbjct: 960 PKSSPPAPVNLPPFEV---KSSPPPTPVSSPPAPK----SSPPAPMSSP-P--PPEV 1009

Query: 176 EPRPPQSPASTASFIFSKGRKQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKS 235
+ PP +P S+ + P P ++ V+ + PP AP S

Sbjct: 1010 KSPPPAPVSSPPPVKSSPPAPVSSP-PPPVKSSPPAPVSSPPPVKSSPPAPISS 1068

Query: 236 PKAPPPVARKPS---VGVPASPSPYRAEPLTAPP 268
P PPPV P V PPP S P P+++PP

Sbjct: 1069 P--PPPVKSSPPAPVSSPPPVKSSPPAPVSSPP 1102

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
Identities = 82/267 (30%), Positives = 110/267 (41%)

Query: 17 PEPAG-PSGSELVSSPAASS---SSATALQIQPGSPDPPAP---PAPAPASSAPGHV 69
P P G P SP + PAAS+ S T + P P+P PP P P P +P

Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSHTPPDVSPPELPEPSVPVAPAPMPMTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGPPREDVGAPLVTPSLQMVRLRSVGA 128
+P PV G S P V P + +V+L AP G+P P + ++P P

Sbjct: 469 DYVPPTTPVPGKSPATSPSPQVQPPAASPPPSLVKLSPPQAPVGSPPPVKTTSPPAP 528

Query: 129 LRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSPASTAS 188
+ G SP P P S + +K+ A G + P PPE P PP AS

Sbjct: 529 I-----GSPSP-PPPVSVVSPPPPVKSSPPAPVGSPP--PPEKSSPPAPVASPP 577

Query: 189 FIFSKGRKQLERPV---SPETQADLQRNLVAELRS-----ISEQRPPQA-----PK 233
+ S L P SP A + + ++S ++ PP P

Sbjct: 578 PVKSSPPPTLVASPPPVKSSPPAPVA-SPPPVKSSPPPTPVASPPPPAPVASSPPPM 636

Query: 234 KSPKAPPPVARKP---SVGVPASPSPYRAEPLTAPPTN 270
KSP P PV+ P PPP + S P E PPT+

Sbjct: 637 KSPPPPTPVSSPPPEKSSPPPPPAKSTPPPEEYPTPTS 676

Score = 170 (25.5 bits), Expect = 1.6e-09, P = 1.6e-09
Identities = 78/279 (27%), Positives = 108/279 (38%)

Query: 5 PPPEEAFSSVASPEPAGPSGSPSELVSSPAASSSSATALQIQPPGSPDPPPPAPPAPAPASS 64
 PP S S + P +P + P SS A+ PP +P +PP P SS
 Sbjct: 883 PPAPTTVISPPSEPKSSPPPTVSLPPPIVKSSPPFAMVSSPPMTPKS--SPP-PVVVSS 939

Query: 65 APGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPG--GAPTALGP 122
 P V P PV PP +P P L ++S P +P PA
 Sbjct: 940 PPPTVKSSPPAPVS-----SPPATPKSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKS 994

Query: 123 SAPQKPLRRALSG--RASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPP 180
 S P P+ ++ P PAP S V+ S +SS P PP + PP
 Sbjct: 995 SPPAPMSSPPPEVKSPPPAPVSSPPPVK---SPPPPAPVSS--P--PPPVKSPPP 1046

Query: 181 QSPASTASFIFSGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPQAPKKSPPKAPP 240
 +P S+ + P P ++ V+ + PP AP SP PP
 Sbjct: 1047 PAPVSSPPPVKSPPPAPVSSPPPVKSPPPAPVSSPPPVKSPPPAPVSSP--PP 1103

Query: 241 PVARKPS---VGVPFPAS---PSYPRAEPLTAPPTNGLPHTQDRTKREL 283
 P+ P V PPPA PS P P+++PP P + ++ L
 Sbjct: 1104 PIKSPPPAPVSSPPAPVKPSLPPAPVSSPPPVVTPAPPKKEEQSL 1152

Score = 169 (25.4 bits), Expect = 2.1e-09, P = 2.1e-09
 Identities = 75/266 (28%), Positives = 104/266 (39%)

Query: 3 DFPPEEAFSSVASPEPAGPSGSPSELVSSPAASSSSATALQIQPP----GSPDPP---PA 55
 D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +
 Sbjct: 469 DYVPPTTP---VPGKSPPATSPSPQ-VQPPAASPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGA 115
 PPAP + S P V+ + PV PP VG+P P V +P
 Sbjct: 525 PPAPIGSPSPPPVSVVSPPPPVKSP----PPAPVGSPP--PPPEKSPPPAPVASP--- 575

Query: 116 PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEA 175
 P P P P ++ P PAP + V+ S ++S P P +
 Sbjct: 576 PPPVKSPPPPTLVASPPPVKSPPPAPVASPPPVK---SPPPTPVASPPPPAPVAS 631

Query: 176 EPRPQSPASTASFIFSGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPQAPKKS 235
 P P +SP K P P S+ PP+
 Sbjct: 632 SPPPMKSPPPPTPVSSPPPEKSP--PPPPAKSTPPPEEYPTPTSVKSSPPPEKSLPP 689

Query: 236 PK---APPPVARK--PSVGVPFPASPSYPRA--EPLTAPP 268
 P +PPP + PS PP+SP P EP+++PP
 Sbjct: 690 PTLIPSPPPQEKPTPSTPSKPPSSPEKPSPPKEPVSSPP 729

Score = 168 (25.2 bits), Expect = 2.7e-09, P = 2.7e-09
 Identities = 75/267 (28%), Positives = 102/267 (38%)

Query: 2 ADFPPEEAFSSVASPE-PAGPSGSPSELVSSPAASSSSATALQIQPPGSPDPP-PAPPAP 59
 A PPP + ++ P+ P G P +SP A S + SP PP +PP P
 Sbjct: 496 ASTPPP--SLVKLSPPQAPVGSPPPPVKTTSPPAPIGSPSPPPVSVVSPPPPVKSPPPP 553

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTPA 119
 AP S P P PV PP + P + S V+ AP +P P
 Sbjct: 554 APVGSPPPEKSPPPAPVASPP---PPVKSPPPPTLVASPPPVKSPPPAPVASPPPP 610

Query: 120 LGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSL-AAASEGLSSAQPNGPPEAEPR 178
 + P P+ + P PAP + ++ +S P PP A+
 Sbjct: 611 VKSPPPPTVA-----SPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPPAKST 664

Query: 179 PP--QSPASTASFIFSGSRKLQLERPV---SPETQADLQRNLVAELRSISEQRPQAPK 233
 PP + P S S K L P SP Q S ++P +P
 Sbjct: 665 PPPEEYPTPTSVKSSPPPEK-SLPPPTLIPSPPPQEKPTPSTPSKPPSSPEKP--SPP 721

Query: 234 KSPKAPPPVARKPSVGVPFPASPSYPRAEPLTAPP 268
 K P + PP K S PPPA S P P+++PP
 Sbjct: 722 KEPVSSPPQTPKSS---PPFAPVSSPPPTPVSSPP 753

Score = 166 (24.9 bits), Expect = 4.6e-09, P = 4.6e-09
 Identities = 81/268 (30%), Positives = 108/268 (40%)

Query: 5 PPPEEAF---FSVASPEPAGPSGSP-LSVSSPAASSSS---ATALQIQPPGSPDPPPP-- 54
 PPPE++ VASP P S P LV+SP S A PP PPP
 Sbjct: 560 PPPEKSPPPAPVASPPPPVKSPPPPTLVASPPPVKSPPPAPVASPPPVKSPPPPTP 619

Query: 55 --APPAPAPASSAPGHVAKLPQKEPVGC---SKGGGPPREDVGAPLVTPLSLQMVRLRS 108
 +PP PAP +S+P + P PV K PP P ++S
 Sbjct: 620 VASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKS 679

Query: 109 VGAPGGA-PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSA 167
 P + P P L PS P P + + ++P PSS + + S SS
 Sbjct: 680 SPPPEKSLPPPTLIPSP--PQEKPTTPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSP 736

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.